

Cherry Picking and Lemon Dropping to Meet Quality Standards

Alex Chan

Stanford University

June 9, 2021

Introduction

- ▶ In the US, substantial heterogeneity in patient outcomes across organ transplant programs
 - ▶ Underlying drivers are unclear: selection vs. “true” differences in quality
 - ▶ Motivates government regulations aimed at penalizing low quality programs
- ▶ This paper: empirically analyze 2007 introduction of regulatory penalties for poor quality
 - ▶ The raw data suggests improvement in patient outcomes
 - ▶ However, improved patient outcomes could arise in two ways:
 - ▶ **Good**: improve quality, judiciously match organs with patients
 - ▶ **Bad**: selectively transplanting healthier patients and withholding surgeries: *“cherry picking and lemon dropping”*

Table of Contents

Institutional Context

Data

Empirical Strategy

Results

Conclusions

Organ Transplantation in the US

- ▶ Removal of healthy organ from one person and transplantation into another person with organ failure
- ▶ Large unmet demand for transplantation:
 - ▶ Over 110,000 Americans waiting for an organ transplantation
 - ▶ Over 6,000 died waiting for an organ in 2018
- ▶ Outside options: None or costly dialysis (in the case of kidneys)
- ▶ Drop in access across the country without quality improvement is welfare loss

Regulatory Enforcement

- ▶ **Penalty for poor quality, including decertification:**
 - ▶ Before 2007: Centers for Medicare and Medicaid Services (CMS) has no formal certification or participation (for reimbursements) requirements based on quality metrics
 - ▶ CMS introduce regulatory penalties for poor quality in 2007
 - ▶ Transplant programs regulated through performance metrics based on post-transplant 1-year patient and graft survival
- ▶ **Poor quality defined by high observed (O) 1-year deaths or 1-year organ failures relative to expected (E) (next slide)**
 - ▶ Expected outcomes are calculated by the Scientific Registry of Transplant Recipients using a range of risk adjustors

Consequences for Higher-than-expected Deaths or Organ Failure under Regulatory Enforcement

- ▶ Noncompliance “flag”: Observed (O) deaths or graft failures significantly and meaningfully higher than the risk-adjusted expected (E) numbers at the transplant program level:
 - ▶ $O/E > 1.5$
 - ▶ $O - E > 3$
 - ▶ O and E significantly different ($p - value < 0.05$)
- ▶ “Flagged” for noncompliance - necessary but insufficient condition for penalties
- ▶ Three things can happen (based on whether CMS determine that the “flag” suggest systematic issues):
 - ▶ Approval after “consideration of mitigating factors”
 - ▶ Imposition of corrective actions: Systems Improvement Agreement (SIA) programs and disruptive peer reviews
 - ▶ Mandatory shut-down

Information from Report Cards and Expectation of Regulatory Actions

- ▶ Transplant centers receive new information *every 6 months* on observed-vs-expected performance
 - ▶ Program Specific Reports publish semi-annually
 - ▶ Reports actual and expected outcomes and “flags”
 - ▶ Hard to completely anticipate as expected outcomes not known beforehand
- ▶ Transplant programs update expected risk of CMS sanctions
 - ▶ A “second strike” or other red flags compels program to revise belief on whether it will be punished by the CMS
 - ▶ Programs use past flags to predict if escalation beyond actions more severe than consideration of mitigating factors likely

Table of Contents

Institutional Context

Data

Empirical Strategy

Results

Conclusions

Data

- ▶ Organ Procurement and Transplantation Network standard analysis file
 - ▶ *Universe of all solid organ transplant candidates, donors, and recipients* in the United States from 1987 to 2019
- ▶ Transplant-program-level report cards
 - ▶ Observed and expected survival rates and compliance with survival standards
 - ▶ From the Scientific Registry for Transplant Recipients
- ▶ *Actual records of action taken by CMS* **due to noncompliance**
 - ▶ Timing of actions and their resolution
 - ▶ U.S. Department of Health and Human Services (Freedom of Information Act request)

Sample Construction and Summary Statistics

- ▶ **Data organized into six-month periods at the transplant program level**
 - ▶ Consistent with 6-month timing when new report cards come out
 - ▶ **37 periods** starting from 2001
- ▶ **Adult kidney and liver transplants only**
 - ▶ Account for most (80%) of the of transplants
 - ▶ Large # of transplant programs (**482**)
- ▶ **CMS took action 169 times**
 - ▶ Impacted 80 kidney/liver programs
 - ▶ 49 kidney/liver cases not approved under mitigating factors and further **escalated into imposition of corrective actions (SIA) or mandatory shut-down**

Main Outcomes of Interest

- ▶ **Quality**
 - ▶ 1-year post transplant patient mortality
 - ▶ 1-year post transplant graft failure
- ▶ **Composition of patients (latent health characteristics)**
 - ▶ Predicted 1-year post transplant patient mortality and 1-year post transplant graft failure using only predictors *included in risk adjustment* in report cards
 - ▶ Coronary Artery Disease and Angina diagnoses, health indicators *NOT included in risk adjustment* in report cards
- ▶ **Quality of organs accepted/declined**
 - ▶ Donor Risk Index (DRI)
 - ▶ Organs from Extended Criteria Donors (older/sicker)
- ▶ **Access**
 - ▶ Number of transplants
 - ▶ Number of newly added candidates

Table of Contents

Institutional Context

Data

Empirical Strategy

Results

Conclusions

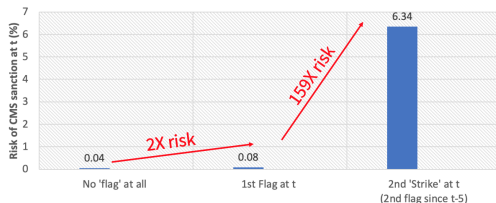
Empirical Strategy: Difference-in-differences

$$Outcome_{it} = \beta T_{it} \cdot \mathbf{1}(t \geq Jul2007) + \alpha_1 T_{it} + \alpha_i + \alpha_t + X'_{it}\gamma + \epsilon_{it}$$

- ▶ An observation is a transplant program(i)-6-months(t) combination
- ▶ T_{it} : treatment variable, indicator of whether a program i has high expected risk of CMS sanctions at t (more on next slides)
- ▶ α_i : time invariant transplant program characteristics. α_t : 6-month fixed effects. X_{it} : risk adjustment (expected 1-year outcomes).

When do program shift practices in light of risk of sanctions?

- ▶ Surgeon interviews: A program will respond to potential CMS sanctions only if the risk of sanctions is material relative to baseline



- ▶ This “treatment” status is *fuzzy*:
 - ▶ Forward-looking transplant doctors and program managers (expectation of CMS punitive actions rather than actual punishment)
- ▶ Use a better predictor of possible sanction than a single flag
 - ▶ Expected risk for CMS punitive actions at t nonparametrically predicted using noncompliance flags from $t - 6$ to t
 - ▶ 95th %-tile as “high risk”

Table of Contents

Institutional Context

Data

Empirical Strategy

Results

Conclusions

Impact on 1-Year Post-Transplant Mortality

- ▶ 2.38 p.p. drop in raw mortality post 2007 but di-in-di estimates are not significantly different from zero
- ▶ Cannot reject hypothesis of no actual quality improvement at all conventional confidence levels (90%+)

Table 5: Impact of quality enforcement by CMS on 1-year patient mortality (target metric)

	(1)	(2)	(3)	(4)	(5)
			1-Year Patient Mortality		
High Risk*Post		-0.00682 (0.00831)	-0.00688 (0.00829)	-0.00602 (0.00831)	-0.00608 (0.00829)
High Risk		0.00341 (0.00730)	0.00334 (0.00727)	0.00384 (0.00727)	0.00376 (0.00723)
Program fixed effects	Y	Y	Y	Y	Y
Time fixed effects	N	Y	Y	Y	Y
SRTR risk adj.	Y	Y	Y	Y	Y
CAD, Angina (≠ risk adj.)	N	N	Y	N	Y
Organ risk (DRI, ECD)	N	N	N	Y	Y
N	11627	11627	11627	11627	11627
R ²	0.259	0.279	0.280	0.283	0.283

Standard errors in parentheses

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

▶ Regression Table for 2nd Strike

▶ Regression Table for Live Donors

▶ Regression Table for Live Donors for 2nd Strike

Impact on 1-Year Post-Transplant Graft Failure

- ▶ 3.26 p.p. drop in raw graft failure rate post 2007 but difference-in-difference estimates are not significantly different from zero
- ▶ Cannot reject hypothesis of no actual quality improvement at all conventional confidence levels (90%+)

Table 6: Impact of quality enforcement by CMS on 1-year graft failure (target metric)

	(1)	(2)	(3)	(4)	(5)
	1-Year graft failure				
High Risk*Post		-0.00714 (0.00689)	-0.00709 (0.00689)	-0.00620 (0.00677)	-0.00614 (0.00677)
High Risk		-0.00139 (0.00631)	-0.00170 (0.00632)	-0.000901 (0.00618)	-0.00122 (0.00620)
Program fixed effects	Y	Y	Y	Y	Y
Time fixed effects	N	Y	Y	Y	Y
SRTR risk adj.	Y	Y	Y	Y	Y
CAD, Angina (\neq risk adj.)	N	N	Y	N	Y
Organ risk (DRI, ECD)	N	N	N	Y	Y
<i>N</i>	11627	11627	11627	11627	11627
<i>R</i> ²	0.150	0.173	0.174	0.179	0.179

Standard errors in parentheses

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

▶ Regression Table for 2nd Strike

▶ Regression Table for Live Donors

▶ Regression Table for Live Donors for 2nd Strike

Effects of Quality Enforcement on Tx Patients Composition

- ▶ Dependent variables below are patient latent health indicators that are not adjusted for performance metrics (CAD, Angina) or risk adjusted (others)
- ▶ The signs of the coefficients are consistent with selection but not significantly different from zero
- ▶ Cannot reject hypothesis of no patient selection at all conventional confidence levels (90%+)

Table 2: Change in composition in transplant patients in response to quality enforcement

	(1) CAD ∉ risk-adj	(2) Angina ∉ risk-adj	(3) E[Mortality] ∈Risk-adj	(4) E[Graft Fail] ∈Risk-adj	(5) Diabetes ∈Risk-adj	(6) Hypertension ∈Risk-adj
High Risk*Post	-0.00286 (0.0124)	-0.00471 (0.0129)	0.00143 (0.000954)	0.000605 (0.000610)	0.00815 (0.00736)	-0.000201 (0.0121)
High Risk	0.00790 (0.0107)	0.00742 (0.0109)	-0.000605 (0.000694)	-0.000838 (0.000546)	-0.0112 (0.00742)	0.000738 (0.0107)
Program fixed effects	Y	Y	Y	Y	Y	Y
Time fixed effects	Y	Y	Y	Y	Y	Y
SRTR risk adj.	Y	Y	N	N	Y (-DM)	Y (-HTN)
<i>N</i>	11627	11627	11627	11627	11627	11627
<i>R</i> ²	0.375	0.436	0.966	0.975	0.396	0.678

Standard errors in parentheses

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Effects of Quality Enforcement on Access and Organ Selection

- ▶ Access dropped: **transplant volume is lower** for programs that became high risk for sanctions **after 2007** relative to those **before 2007**
- ▶ **Transplanted organs**: organs less likely from a high risk donor (accepted organs have **lower risk (DRI)**)

Table 4: Change in transplant volume and composition in deceased donor organs quality in response to quality enforcement

	(1) # transplanted	(2) Donor Risk Index
High Risk*Post	-3.109* (1.312) ✓ Reduced Access	-0.0248+ (0.0145) ✓ Organ selection
High Risk	0.541 (1.098)	-0.0131 (0.0138)
Program fixed effects	Y	Y
Time fixed effects	Y	Y
SRTR risk adj.	Y	Y
<i>N</i>	11627	11627
<i>R</i> ²	0.747	0.762

Standard errors in parentheses

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Table of Contents

Institutional Context

Data

Empirical Strategy

Results

Conclusions

Discussion, Conclusions, Next Steps

- ▶ Transplant programs reduce access and lemon drop organs in response to high risk of facing quality enforcement actions
- ▶ No net quality improvement
- ▶ Calls into question the effectiveness of quality standard enforcement when quality metrics are imperfect