CASE VOLUME AND MORTALITY: THE EFFECT OF INTERVENING ON PHYSICIANS

Arın L. Madenci *
Harvard T.H. Chan School of Public Health
Brigham and Women's Hospital
Boston, MA
arin_madenci@g.harvard.edu

Kerollos Nashat Wanis
Harvard T.H. Chan School of Public Health
Boston, MA

Zara Cooper
Brigham and Women's Hospital
Center for Surgery and Public Health
Boston, MA

SV Subramanian
Harvard T.H. Chan School of Public Health
Boston, MA

Sebastien Haneuse
Harvard T.H. Chan School of Public Health
Boston, MA

Albert Hofman
Harvard T.H. Chan School of Public Health
Boston, MA

Miguel A. Hernán
Harvard T.H. Chan School of Public Health
Boston, MA

August 10, 2020

*Corresponding author

ABSTRACT

Background: Researchers and policymakers seeking to optimize healthcare delivery often consider whether medical services should be regionalized to expert providers or specialized centers. On the one hand, these services would become less readily available to the patients who need them; on the other hand, clinicians may become more familiar and adept with the repetition of frequently treating the same conditions. Research studies have previously identified a correlation between the number of certain operations that surgeons perform and their patients' outcomes; the number of radiographs interpreted by radiologists and their diagnostic performance; and the frequency with which a hospital treats patients with certain conditions and their patients' outcomes. However, previous analyses have often been cross-sectional and have not considered time-varying confounding or common positivity violations.

Methods: We consider an application of this type of research in order to demonstrate common principles and a methodology that may be adapted to numerous other scenarios of substantive interest. Using the example of surgeons who perform coronary artery bypass grafting (CABG) operations for United States Medicare beneficiaries, we make explicit assumptions to evaluate the effect of intervening in different ways on the operative volume of surgeons.

Results: We first describe four hypothetical randomized trials of interest and then demonstrate how to emulate these trials among surgeons who have performed CABG operations for United States Medicare beneficiaries. We draw attention to areas of bias in previously published observational analyses that occur due to deviations from the hypothetical randomized trial of interest. In each target trial emulation, we report that interventions on physician volume of CABG operations were ultimately found to have little effect on patient mortality, except potentially in cases of extreme restructuring of referral patterns.

Conclusions: When the substantive interest lies in intervening on physicians' patient volumes, this manuscript demonstrates how the target trial framework can be used to estimate the resulting effect on the resulting outcomes. This overall methodological approach can be modified for application to other operations and research questions in areas of healthcare outside of surgery.

***K*eywords** Target trial · Health services research · Operative volume · Coronary artery bypass grafting · Inverse Probability Weighting · Marginal Structural Model

Key Messages

- Previous observational data analyses have drawn the conclusion that physicians must achieve a minimum case volume in a given period of time to maintain skills and provide save healthcare services.
- However, these prior studies have often unnecessarily introduced bias or misinterpreted results.
- Taking the example of surgeons who perform coronary artery bypass grafting operations, the authors delineate hypothetical randomized trials which would answer questions related to interventions on physician volume and then demonstrate how to emulate these trials using observational data.

1 Introduction

Policymakers often consider whether medical services should be regionalized to expert providers or specialized centers. Although regionalized services would become less readily available for some patients, the quality of healthcare might increase if clinicians become more adept due to repetition of frequently treating the same conditions.

Researchers have therefore studied the relationship between the frequency with which clinicians perform a specific healthcare task (their "volume" of this task) and the outcomes of their patients. For example, Medicare patients who developed serious conditions had lower mortality when admitted to hospitals that more frequently treated those conditions;[1] radiologists who more frequently interpreted diagnostic mammograms had better performance;[2] and Medicare patients who underwent an operation had lower mortality when the surgeon performed the operation more frequently, compared with hospitals, radiologists, and surgeons, respectively, who less frequently performed these tasks.[3]

The above findings have been interpreted as an indication that increasing clinicians' volume leads to better outcomes. In fact, patient advocacy organizations such as the Leapfrog Group have suggested that healthcare systems should institute minimum-volume standards which withhold credentialing, unless "the hospitals and their surgeons do [the operations] often enough to keep their skill level up."[4] However, the association between increased volume and better outcomes has not been consistently demonstrated for some tasks, such coronary artery bypass graft (CABG).[3,5,6]

The findings from studies of the effect of volume are hard to interpret because they have generally been cross-sectional and have not considered time-varying confounding or common positivity violations. Also, the interpretations of the findings often conflate effects of two different interventions: i.) the effect of *assigning a patient* to select a physician of certain volume and ii.) the effect of *assigning a physician* to a certain volume. That is, the existing body of research does not distinguish between intervening on patients and intervening on physicians (or hospitals).

Here we describe a methodology to quantify the effect of intervening on surgeons' operative volume. We estimate the effect on mortality of assigning surgeons to a specific CABG volume in four hypothetical randomized trials ("target trials") of increasing complexity. We then emulate each of these target trials using observational data from Medicare. This methodology can be adapted to other clinical scenarios to study the effect of increasing or decreasing volume on patient outcomes.

2 Specification of the Target Trials

With unlimited resources and cooperation, it would be hypothetically possible to enroll cardiothoracic surgeons into a randomized trial. We would include surgeons who have performed at least one CABG operation during each of two consecutive intervals of, say, 90 days. We could then assign these surgeons to perform a particular number of operations for patients seeking a first-time CABG operation.

However, it would be unrealistic to simply assign surgeons to an arbitrary volume of operations. Surgeons who perform few operations (for example, 1 or 2), perhaps because they have other obligations, may not be logistically able to suddenly perform 20 operations during the next interval. Therefore, we will consider target trials that compare strategies under which surgeons increase or decrease their operative volume relative to their baseline volume.

For all trials, we consider one or more 90-day intervals during which surgeons must perform their assigned number of operations and a subsequent 90-day interval during which the outcome is evaluated. For each surgeon, the outcome is the proportion of all-cause post-operative mortality among their patients who underwent an operation during the interval subsequent to the completion of the intervention. Trial #1 compares strategies sustained for one interval only. Trial #2 compares strategies sustained over several intervals. Target Trials #3 and #4 are modifications of #1 and #2 that do not assume the existence of an unlimited number of available patients.

2.1 Target Trial #1: Single interval intervention

We randomly assign each eligible surgeon i to perform a number of CABG operations during the baseline interval $k = 0$ that is equal to the surgeon's operative volume during the pre-baseline interval $k = -1$ plus/minus a random number x that can take values in $\{-5, -4, \ldots, -4, 5\}$. This 11-arm trial is represented by the causal directed acyclic graph (DAG) in Figure 1.

2.2 Target Trial #2: Sustained intervention

We randomly assign each eligible surgeon to perform a number of CABG operations per interval for four consecutive intervals ($k \in \{0, \ldots, K\}$, where $K = 3$) that is equal to the surgeon's operative volume during the pre-baseline interval $k = -1$ plus a random number x that can take values in $\{-5, -4, \ldots, -4, 5\}$. To allow for vacations, travel to conferences, or patient cancellations, we stipulate that surgeons adhere to their exact assignment for at least three of the four 90-day intervals. This 11-arm trial is represented by the causal DAG in Figure 2.

2.3 Target Trials #3 and #4: Modification for real-world constraints

If we were interested in a national policy, then acting on the results of Trials #1 and #2 would assume a sufficiently large number of eligible patients whose date of operation can be modified to accommodate all of the surgeons who were assigned to perform more operations. Were surgical simulation to become realistic enough, we could imagine a future in which surgeons could perform simulated CABG operations to compensate for any shortfall in the operative volume that we assign. However, in the current world, even with limitless financial resources and cooperation, such restructuring of operative volumes would not be feasible.

Trials #3 and #4 address this practical problem by keeping the net number of patients fixed relative to the pre-baseline interval. Specifically, Target Trial #3 (and #4) resembles Target Trial #1 (and #2) with the following modification. Surgeons are again assigned to perform a number of CABG operations per interval that is a function of a random number x (ranging from -5 to 5) and their pre-baseline volume. However, in these trials, surgeons who performed *more* operations than the median baseline volume of all surgeons are assigned to change their volume by $-x$ operations, while below-median surgeons are assigned to change their volume by $+x$ operations. This strategy effectively redistributes patients among surgeons. For scenarios requiring below-median surgeons to reduce their volume below zero (i.e., for $x < 0$), we assign them to perform zero operations and randomly redistribute their patients to above-median surgeons (see Appendix Section 8.1 for details).

2.4 Analysis of the target trials

Had the randomized trials described above been implemented, we could nonparametrically estimate the 90-day risk of mortality among operations performed during interval $K + 1$ by each surgeon, π, and then the mean mortality risk in each of the 11 arms indexed by x, $\mathbb{E}\big[\pi|X = x\big]$. In the absence of loss to follow-up (surgeons who do not perform any operations during this $K + 1^{\text{th}}$ interval are considered lost to follow-up.), a contrast of these mean risks quantifies an intention-to-treat effect, that is, the effect of assignment to each volume on post-operative mortality in the k^{th} interval. However, given the large number of strategies relative to the number of included surgeons, we can obtain more precise estimates with decreased variances by making parametric assumptions that smooth across strategies. To do so, we could compute predictions for each arm x from a pooled logistic model:

$$\mathbb{E}\big[\pi|X = x\big] = \text{expit}\big(\alpha_0 + \alpha_1 f(x)\big)$$

where $f(x)$ denotes a flexible functional form of treatment arm (such as restricted cubic splines).

Alternatively, we may be more interested in the per-protocol effect,[7] that is, the effect that would have been observed had all surgeons adhered to their assigned intervention arm and had complete follow-up.

To estimate the mean risk under perfect adherence to each of the strategies indexed by x, we first fit the outcome model

$$\mathbb{E}\big[\pi|X = x, V, W\big] = \mathbb{E}\big[\pi|A = g(X, W), V, W\big] = \text{expit}\big(\alpha_0 + \alpha_1 f(a - W) + \alpha_2^T V\big)$$

where A denotes the number of operations performed during each interval, W denotes operative volume in the pre-baseline interval, $g(X, W) = W + X$, and V denotes the following surgeon-specific baseline covariates: surgeon age, surgeon gender, date of meeting eligibility criteria, and average total number of hospital beds at their hospital(s). For Trial #2, which includes several intervals, we censor surgeons when they deviate from the strategy to which they were randomly assigned at baseline.

We then standardize by the baseline covariates

$$\begin{aligned}\mathbb{E}\big[\pi|X = x\big] &= \sum_{w \in \mathcal{W}, v \in \mathcal{V}} \mathbb{E}\Big[\pi|A = g(x, w), W, V\Big] p(W = w, V = v) \\ &= \hat{\mathbb{E}}_{w \in \mathcal{W}, l \in \mathcal{L}} \mathbb{E}\Big[\pi|A = g(x, w), W, V)\Big]\end{aligned}$$

The unbiased estimation of the per-protocol effect requires covariate adjustment to eliminate potential bias from imperfect adherence and loss to follow-up. The standardized estimates described above adjust for baseline covariates V, which is sufficient for estimating the per-protocol effect after one interval, as in Trial #1.

On the other hand, estimating the per-protocol effect over several intervals as in Trial #2 requires adjustment for time-varying (post-baseline) covariates L_k, including the surgeon's hospital characteristics (i.e., CABG operative volume) and patient characteristics (proportion of patients with a history of acute myocardial infarction, atrial fibrillation, chronic kidney injury, chronic obstructive pulmonary disease, congestive heart failure, diabetes, stroke or transient ischemic attack, and dementia; proportion of patients who underwent elective hospital admission; average hospital proportion of Medicare patients; interval number; and proportion of patients who died within 90-days post-operatively after undergoing an operation during the prior interval (if zero operations were performed in an interval, this proportion was set to zero). To adjust for time-varying covariates that may be associated with past adherence.[8,9] we use (stabilized) inverse probability weights as described in the Appendix.

The analyses for Target Trials #3 and #4 mirror those of Target Trials #1 and #2, respectively.

A non-parametric bootstrap procedure with 1000 resamples can be used to estimate 95% confidence intervals for all trials.

3 Emulation of the Target Trials

We cannot conduct these target trials, because, for example, referring physicians may not agree to have their patients randomly assigned to a surgeon with whom they are not familiar and surgeons may not agree to forfeit the payment associated with randomization to an arm with lower operative volumes. But we can attempt to emulate these target trials using observational data.[10]

Our data source is a 100% sample of fee-for-service Medicare claims from 1 January 2011 to 30 September 2016, filed by individuals who were over 65 years of age and had not previously undergone a CABG operation (Appendix Table 1. Observation units for each surgeon were partitioned into 90-day discrete time intervals. Nationally, 85% of CABG operations among patients age 65 or older were estimated to have been paid by Medicare.[11]

3.1 Eligibility criteria

Surgeons were identified using a unique provider identification number designated by the "primary operator" field of the inpatient Medicare data and subspecialty was classified by the surgeon specialty identifier in the Medicare Data on Provider Practice and Specialty file.[12] Data from the American Hospital Association Annual Survey of Hospitals was used to determine the characteristics of included hospitals. Surgeon and hospital characteristics were obtained from the Medicare Data on Provider Practice and Specialty and the American Hospital Association Annual Survey of Hospitals, respectively. Average patient comorbidity information was obtained from the Medicare Master Beneficiary Summary File.[13]

3.2 Strategies

Each surgeon's 90-day operative volume, defined as the count of CABG operations performed in the Medicare inpatient claims file, was recorded at the end of each interval (i.e., at the beginning of the k^{th} interval, surgeon operative volume was counted over the prior 90-day interval). Each surgeon was assigned to the intervention arm they were observed to have followed.

Operations performed for individuals other than fee-for-service Medicare beneficiaries age 65 or older and undergoing their initial CABG were not included in the count. Therefore, these strategies may underestimate the total operative volume of each surgeon.

3.3 Randomized assignment

We emulated randomization of surgeons to a particular operative volume by adjusting for the following baseline covariates which are believed to be confounders: surgeon age, surgeon gender, date of meeting eligibility criteria, and average total number of hospital beds at their hospital(s).

3.4 Follow-up

The beginning of follow-up starts at the time of meeting eligibility criteria and extends for up to four 90-day intervals, plus 90 days during which the outcomes are measured. A sensitivity analysis for Target Trial #2, presented in the Appendix, required surgeons to perform at least one operation during each interval.

3.5 Outcome

Patient mortality was assessed by linkage with the Medicare Vital Status file, which integrates information from Medicare claims data, family members, and the Railroad Retirement Board and the Social Security Administration.

3.6 Causal contrast

We estimated the observational analog of per-protocol effects.

3.7 Statistical analysis

The emulation of the target trials can be accomplished using the same analyses as described above in Section 2.4, with the following modifications.

In all trials, for statistical efficiency, if a participant meets eligibility criteria more than once, multiple nested target trials will be emulated (each with a different time of baseline, dependent on when the eligibility criteria were met). In the emulation of Target Trials #2 and #4 every surgeon has data compatible with more than one strategy at baseline. In fact, because all strategies allow for one interval during which surgeons are excused from reaching their assigned volume, all individuals have data compatible with all strategies in the first interval ($k = 0$). A statistically efficient way to proceed in face of this complexity is to expand the dataset with "cloned" participants.[14,15] Each participant will thus have as many copies (clones) as regimes they are observed to follow. Each clone is censored at the time of deviation from their assigned treatment strategy. An example of this censoring procedure is presented in Appendix Table 3. Code used for the analysis of this manuscript is available at `github.com/arinmadenci/volume-surgeon`.

4 Estimates from Medicare data

The baseline characteristics of the 2,338 eligible surgeons, their patients, and the hospitals in which they performed the operations are displayed in Table 1.

The mean surgeon operative volume was 5.4 operations per 90-day interval and ranged from <1 to 41. Surgeons who performed an average of 15 operations and 21 operations per 90-day interval were in the top 95$^{\text{th}}$ and 99$^{\text{th}}$ percentile, respectively. Eligible surgeons were observed to perform a change in the number of operations from their baseline volume between -29 and 28 in a single interval or between -29 and 13 sustained for at least three of four consecutive intervals. Because of the higher variance associated with higher operative volume, there were fewer surgeons who had high baseline operative volumes and followed the sustained regimes (compared with those with lower baseline operative volumes). However, there did not appear to be non-random violations of positivity (Appendix Table 5).

The mean surgeon-specific mortality in the $K + 1^{\text{th}}$ interval was 6.0% for both $K = 0$ (Trials #1 and #3) and $K = 3$ (Trials #2 and #4). Mortality estimates for each of the different trials are summarized in Table 2 and Figure 3. We estimated that, had all surgeons increased their CABG operative volume by 5 for one 90-day interval as in Trial #1 (or a full year, as in Trial #2), the expected mortality would subsequently be 5.7% (4.5%), compared with 6.1% (6.9%) had all surgeons instead decreased their volume by 5.

Alternatively, we estimated that, had 5 CABG operations from above-median surgeons been re-assigned to below-median surgeons for one 90-day interval as in Trial #3 (or a full year, as in Trial #4), the expected mortality would be 5.9% (3.7%), compared with 5.8% (6.9%) had 5 operations been re-assigned from below-median surgeons to their above-median volume counterparts.

We performed several sensitivity analyses, modifying the emulation of Target Trials #2 and #4. First, for Trial #2, we repeated the IP weighting procedure, but required that surgeons perform at least one operation in all periods (in the main analysis surgeons were allowed to deviate from the regime in one interval including performing zero operations). These estimates are reported in Appendix Table 6. The approach is described in detail in the Appendix. Additionally, for Target Trials #2 and #4, estimates from each intervention arm x made separately (i.e., without a parametric assumption on the relationship between intervention and outcome) are presented in Appendix Tables 7 and 8.

5 Discussion

We used Medicare data to emulate four target trials of changes in surgeon's CABG operative volume. Single-interval interventions on operative volume (Trials #1 and #3) had minimal effect on patient mortality. Likewise, we found little effect on patient mortality of interventions on operative volume that were sustained over a calendar year (Trials #2 and

#4), except potentially in cases of more extreme restructuring of referral patterns. However, the confidence intervals were wide for such extreme interventions.

Previous publications studying the effect of volume on patient outcomes have yielded inconsistent conclusions. Some studies have suggested that increasing physicians' volume would substantially improve patient outcomes,[1,3], while another has emphasized the complexities of this relationship.[2] Studies specific to CABG operations reported conflicting results;[3,5,6] however, these analyses were cross-sectional and applied methodologies not suitable to study interventions on physicians (as opposed to patients). Given that these analysis plans additionally did not consider positivity violations, time-varying confounding, and real-world constraints on implementation, their estimates are hard to interpret causally and act on.

In this manuscript, we describe target trials that explicitly evaluate intervening on case referrals to surgeons. We explain why assigning surgeons to a volume regardless of their baseline volume is unrealistic. We first consider a single interval intervention trial to introduce key concepts without distraction from the complexity of inverse probability weights for interventions sustained over more than one interval. Prior studies have not considered the constraints of restructuring due to a finite number of available patients. Therefore, we initially considered trials with interventions that assumed a limitless number of patients. These single and multiple interval analyses which emulate trials that increase or decrease baseline volumes were more reasonable than prior trials that ignored baseline volume; however, this type of study does not account for real-world constraints in the number of patients available for assignment. The final, and most realistic, type of trial relaxes this assumption by directly incorporating interventions which reassign cases from lower volume to higher volume surgeons (and *vice versa*). Due to limitations in the the observed data, we did not emulate trials in which the intervention was sustained for many years.

For researchers and policymakers interested in intervening on physician case volume, these methods may help to clarify several analytical complexities. First, it is important to distinguish between interventions on physicians (or hospitals) and interventions on patients. Second, for interventions on physicians, there are practical issue to consider of i.) unrealistic interventions specifying very low volume physicians begin to perform many cases despite other responsibilities and ii.) a finite number of patients such that it is impossible to restructure referral patterns by requiring all physicians to increase their baseline volume by a certain number of cases. Furthermore, certain specifics of the analytical approach will likely differ depending on the procedure or task of interest. Emulating more realistic target trials (similar to Trial #4) may provide policymakers with more actionable estimates.

In summary, we outlined and applied a methodology that estimates the expected outcome of requiring a surgeon to perform a certain volume of CABG operations. When health services research involves substantive interest in intervening on physicians' patient volumes, this manuscript demonstrates how the target trial framework can be used to estimate the resulting effect on the resulting outcomes. This overall methodological approach can be modified for application to other operations and non-surgical questions.

6 Figures

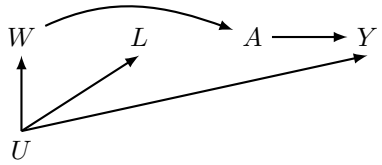

Figure 1: Directed acyclic graph corresponding to target trial #1. W denotes surgeon operative volume from interval $k = -1$; L is a random vector of covariates from interval $k = -1$, including surgeon characteristics, average patient characteristics, and average hospital characteristics; A is the assigned operative volume to be performed during the interval $k = 0$; Y is surgeon-specific proportion of mortality of patients who underwent an operation during interval $k = 1$; and U is a random vector of unmeasured covariates. Directed edges between W, L, and Y are omitted for simplification.

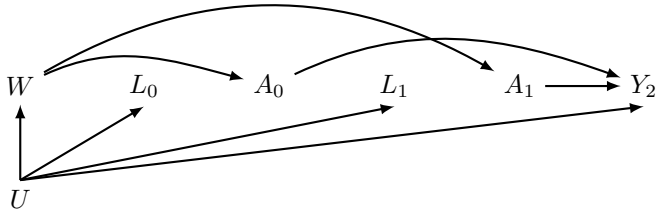

Figure 2: Directed acyclic graph corresponding to target trial #2. W denotes surgeon volume during interval $k = -1$; L_k is a random vector of covariates from interval k, including surgeon characteristics, average patient characteristics (including past outcomes), and average hospital characteristics (L_0 also includes V, denoting the random vector of time-fixed surgeon and hospital covariates); A_k is a random variable denoting the number of operations performed by a given surgeon in the interval k; Y_k is mortality of patients who underwent an operation during the k^{th} interval; and U is a random vector of unmeasured covariates. Directed edges between W, L_k, and Y are omitted for simplification.

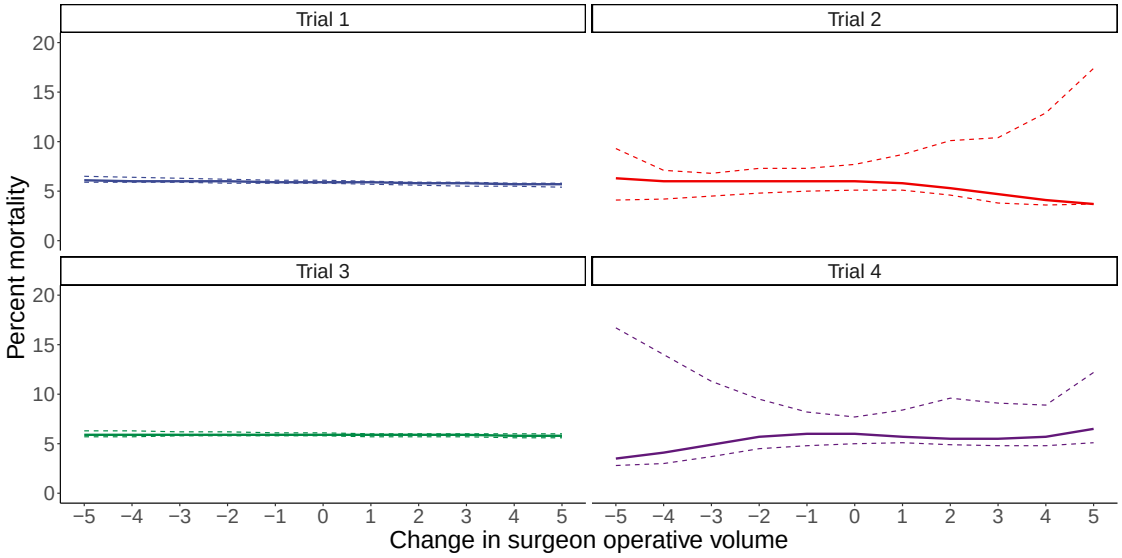

Figure 3: Ninety-day mortality estimates corresponding to the emulation of each target trial.

7 Tables

Table 1: Baseline characteristics (during the previous 6 months) of 2338 eligible surgeons who performed CABG for U.S. Medicare beneficiaries

	Number (%) or mean (s.d.)
Surgeon characteristics	
Mean count of CABG operations per 90 days	8.1 (6.3)
Mean surgeon age, years	50.6 (9.0)
Number of female surgeons	86 (3.7)
Hospital characteristics	
Total number of hospitals	1035
Mean hospital volume (operations per 90-day interval)	29.2 (24.1)
Mean hospital number of beds	506.1 (324.7)
Hospital proportion of patients with Medicare	0.4 (0.1)
Case mix characteristics	
Total number of patients	37991
Mean patient age, years	74.2 (3.2)
Proportion of patients with AMI	0.1 (0.2)
Proportion of patients with dementia	0.1 (0.1)
Proportion of patients with atrial fibrillation	0.2 (0.2)
Proportion of patients with CKD	0.3 (0.2)
Proportion of patients with COPD	0.3 (0.2)
Proportion of patients with CHF	0.4 (0.3)
Proportion of patients with diabetes	0.5 (0.3)
Proportion of patients with stroke or TIA	0.2 (0.2)

Table 2: Mortality estimates (%) for emulation of Target Trials

Assignment x	Trial #1	Trial #2	Trial #3	Trial #4
-5	6.1 (5.9-6.5)	6.3 (4.1-9.3)	5.9 (5.7-6.3)	3.5 (2.8-16.7)
-4	6.0 (5.9-6.4)	6.0 (4.2-7.1)	5.9 (5.7-6.3)	4.1 (3.0-14.0)
-3	6.0 (5.9-6.3)	6.0 (4.5-6.8)	5.9 (5.8-6.2)	4.9 (3.7-11.3)
-2	6.0 (5.8-6.2)	6.0 (4.8-7.3)	5.9 (5.8-6.2)	5.7 (4.5-9.5)
-1	5.9 (5.8-6.1)	6.0 (5.0-7.3)	5.9 (5.8-6.1)	6.0 (4.8-8.2)
0	5.9 (5.8-6.1)	6.0 (5.1-7.7)	5.9 (5.8-6.1)	6.0 (5.0-7.7)
1	5.9 (5.7-6.0)	5.8 (5.1-8.7)	5.9 (5.7-6.0)	5.7 (5.1-8.4)
2	5.8 (5.6-5.9)	5.3 (4.6-10.1)	5.9 (5.7-6.0)	5.5 (4.9-9.6)
3	5.8 (5.5-5.9)	4.7 (3.8-10.4)	5.9 (5.7-6.0)	5.5 (4.8-9.1)
4	5.7 (5.5-5.8)	4.1 (3.6-12.9)	5.8 (5.6-6.0)	5.7 (4.8-8.9)
5	5.7 (5.4-5.8)	3.7 (3.7-17.4)	5.8 (5.6-6.0)	6.5 (5.1-12.2)

In Trial #1 (or Trial #2), x denotes an addition of CABG operations for one (or four) 90-day interval(s). In Trial #3 (or Trial #4), x denotes an addition of CABG operations for below-baseline surgeons and a subtraction for above-baseline surgeons for one (or four) 90-day interval(s).

8 Appendix

8.1 Appendix: Assignment mechanism for Target Trials #3 and #4

Surgeons who performed fewer operations than the median volume of all surgeons during the pre-baseline interval (denoted by $\tilde{\mu}$) are assigned to add to their volume by x cases (as in prior trials). For many surgeons with pre-baseline volume $w < \tilde{\mu}$, they will not be able to fully add to their volume when $x < -1$ and $|x| > w$; as such, each surgeon j is assigned to perform $A_j = \max(0, w_j + x)$ as in trials #1 and #2. For surgeons with $w \geq \tilde{\mu}$ and $x \geq -1$, they are assigned to perform $w - x$ operations. However, for surgeons with $w \geq \tilde{\mu}$ and $x < -1$, fewer than x cases will be available for each surgeon; specifically, only $\sum_{j:w_j<\tilde{\mu}} w_j - a_j < x \sum_{j=0}^{J} \mathbb{1}(w_j < \tilde{\mu})$ cases will be available. As such, when $x < -1$, surgeons with $w \geq \tilde{\mu}$ are randomly assigned to a volume $w_j - a_j$ for a randomly selected surgeon j with $w < \tilde{\mu}$ such that the net number of operations will remain fixed at the pre-baseline level.

In summary, for intervention arm x and pre-baseline volume $w_{(1)}, \ldots, w_{(J)}$ in ascending order with surgeons $j \in \{1, \ldots, J\}$, each surgeon's operative volume assignment will be

$$A_j = \begin{cases} \max(w_j + x, 0) & \text{if } \; w_j < \tilde{\mu} \\ \max(w_j - x, 0) & \text{if } \; w_j \geq \tilde{\mu} \; \text{ and } \; x \geq -1 \\ n_m & \text{if } \; w_j \geq \tilde{\mu} \; \text{ and } \; x < -1 \end{cases} \tag{1}$$

where n_m for $m \in \{\lceil J/2 \rceil + 1, \ldots, J\}$ is defined by the following:

Let $\mathbf{k} = k_1, \ldots, k_{\lceil J/2 \rceil} = w_1 - a_1, \ldots, w_{\lceil J/2 \rceil} - a_{\lceil J/2 \rceil}$ and $\mathbf{k}^*$ be a random permutation of $\mathbf{k}$: $k_1^* = k_{\pi(1)}, \ldots, k_{\lceil J/2 \rceil}^* = k_{\pi(\lceil J/2 \rceil)}$.

Then define $\mathbf{n} = \mathbf{k}^* = k_{\pi(1)}^*, \ldots, k_{\pi(\lceil J/2 \rceil)}^*$.

8.2 Appendix: Consideration of positivity violations

In the Medicare data, eligible surgeons were observed to perform between 0 and 59 operations per interval. Among surgeons with a pre-baseline operative volume of 0-4, very few were observed to perform higher numbers of operations in the following interval. Conversely, among surgeons with a pre-baseline operative volume of 15 or greater, very few were observed to perform lower numbers of operations in the following interval. A summary of observed operative volumes for different pre-baseline operative volume histories is reported in Appendix Table 2. Given this concern for non-random violations of positivity in the available observed data (especially when including other covariate patterns), we did not consider this trial further. These problems were magnified with a static sustained regime.

8.3 Appendix: Inverse probability weights

This section describes the estimation of stabilized inverse probability weights used in the main analysis:

$$SW = SW^A \cdot SW^C$$

$$SW^A = \prod_{k=0}^{K} \frac{f(A_k|A_{k-1}, V)}{f(A_k|A_{k-1}, \overline{L}_{k-1}, V)}$$

where $f(A_k|A_{k-1}, \overline{L}_{k-1}, V)$ is the conditional probability mass function $f_{A_k|A_{k-1},\overline{L}_{k-1},V}(a_k|a_{k-1}, \bar{l}_{k-1}, v)$ evaluated at the random expression $A_k|A_{k-1}, \overline{L}_{k-1}, V$ and, as described in the main text, A is the sum of baseline volume W and assignment X.

Additionally, because a one-interval deviation is permitted, the numerator and denominator are deterministically equal to 1 in the first interval $k = 0$ following consecutive adherence to the randomized assignment (all individuals are by convention considered to be adherent during interval $k = -1$).

To estimate the quantities in the numerator and denominator, under the assumption that the operative volume of a surgeon in any given interval (A_k) is a negative binomial random variable with non-constant dispersion, we fit the following discrete-time regressions:

$$\hat{\mathbb{E}}[A_k|A_{k-1}, V] = \exp\left(\theta_{0,k} + \theta_1^T h(a_k) + \theta_2^T V\right) \tag{2}$$

$$\hat{\mathbb{E}}[A_k|A_{k-1}, \overline{L}_{t-1}, V] = \exp\left(\theta_{0,k} + \theta_1^T h(a_k) + \theta_3^T \bar{l}_k + \theta_4 V\right) \tag{3}$$

where $\theta_{.,k} = \theta_.^T g(k)$ is a time-varying parameter and the overbar denotes a variable's history in the prior interval. The dispersions for Equations (2) and (3) are separately modeled using the same covariates, using gamma regression. Surgeon-specific mean and dispersion predictions for the numerator and denominator equations are then used to generate negative binomial distributions with which corresponding probabilities can be computed for each surgeon-specific observed value of A_k.

Finally, only surgeons who perform one or more operations during interval $K+1$ will have measurable outcomes during that interval. To account for this potential selection bias, the following stabilized censoring weights can be used:

$$SW^C = \frac{p(C_{K+1} = 0|A_K, V)}{p(C_{K+1} = 0|A_K, \overline{L}_K, V)}$$

where C_{K+1} denotes censoring in interval $K+1$. The numerator and denominator of the weights can be estimated with the following discrete-time logistic regressions:

$$\hat{p}(C_{K+1} = 0|A_K = a_K, V = v) = \text{expit}\big(\theta_0 + \theta_1^T h(a_k) + \theta_2^T v\big)$$

$$\hat{p}(C_{K+1} = 0|A_K = a_K, \overline{L}_K = \bar{l}_K, V = v) = \text{expit}\big(\theta_0 + \theta_1^T h(a_k) + \theta_2^T \bar{l}_t + \theta_4 v\big)$$

8.4 Appendix: Sensitivity analysis

This section describes the estimation of stabilized inverse probability weights used in the main analysis:

$$SW = SW_K^A \cdot SW_{K+1}^C$$

8.4.1 Treatment weights

$$SW_t^A = \prod_{k=1}^{t} \frac{f(A_k|A_{k-1}, V, C_k = 0)}{f(A_k|A_{k-1}, \overline{L}_{k-1}, V, C_k = 0)} \tag{4}$$

We fit the same discrete-time regressions for the numerator and denominator as in the main analysis, now conditional on not having been censored ($\overline{C}_k = 0$) for deviating from the assignment a second time or performing zero operations in period k.

$$\mathbb{E}[A_k|A_{k-1}, V, \overline{C}_k = 0] = \exp\left(\theta_{0,k} + \theta_1^T g(A_{k-1}) + \theta_2^T V\right) \tag{5}$$

$$\mathbb{E}[A_k|A_{k-1}, V, \overline{L}_{k-1}, \overline{C}_k = 0] = \exp\left(\theta_{0,k} + \theta_1^T g(A_{k-1}) + \theta_2^T \overline{L}_k + \theta_3^T V\right) \tag{6}$$

where $\theta_{.,k} = \theta_.^T g(k)$ is a time-varying parameter, $g(\cdot)$ is a flexible function such as restricted cubic splines, and the overbar denotes a variable's history including the previous interval.

8.4.2 Censoring weights

After following the regime in each interval k, only surgeons who perform one or more operations during interval $k+1$ will have measurable outcomes during that interval. To account for this censoring, the following weights were used:

$$SW_t^C = \prod_{k=1}^{t} \frac{f(C_k|A_k, V, C_{k-1} = 0)}{f(C_k|A_k, \overline{L}_k, V, C_{k-1} = 0))} \tag{7}$$

The numerator and denominator of the surgeon-specific censoring weights were estimated with the following discrete-time logistic regression equations:

$$\mathbb{E}[C_k|V] = \exp\left(\theta_{0,k} + \theta_1 A_{k-1} + \theta_2^T V\right) \tag{8}$$

$$\mathbb{E}[C_k|A_{k-1}, \overline{L}_{k-1}, V] = \exp\left(\theta_{0,k} + \theta_1 A_{k-1} + \theta_2^T \overline{L}_{k-1} + \theta_4^T V\right) \tag{9}$$

8.4.3 Standardization

As described above, the time-fixed covariates were included in the numerators of the stabilized weights. These variables can be adjusted for using the following outcome regression, weighted by SW:

$$\mathbb{E}\big[\pi|X=x\big] = \sum_{w\in W, v\in V} \mathbb{E}\Big(\pi|X=x,W,V\Big)p(W=w,V=v) \tag{10}$$

$$= \hat{\mathbb{E}}_{w\in W,v\in V}\mathbb{E}\Big[\pi|X=x,W,V)\Big] \tag{11}$$

$$= \hat{\mathbb{E}}_{w\in W,v\in V}\text{expit}\big(\alpha_0+\alpha_1 g(X)+\alpha_2^T V+\alpha_3^T W\big) \tag{12}$$

Interaction terms were additionally included between the intervention arm and all time-fixed covariates.

8.5 Appendix: Tables

Appendix Table 1: International Classification of Diseases (ICD) Codes

ICD code	Revision version	Description
361	ICD-9	Bypass anastomosis for heart revascularization
362	ICD-9	Heart revascularization by arterial implant
02100	ICD-10	Bypass coronary artery, one artery
02110	ICD-10	Bypass coronary artery, two arteries
02120	ICD-10	Bypass coronary artery, three arteries
02130	ICD-10	Bypass coronary artery, four or more arteries

Appendix Table 2: Positivity check for Target Trial #1

Volume category	History: 0 to 4	History: 5 to 9	History: 10 to 14	History: 15+
0	1356	356	543	28
1	1451	392	667	5
2	1632	750	1140	5
3	1575	1061	1510	13
4	1418	1298	1830	32
5	1132	1436	1976	74
6	857	1479	1939	76
7	603	1415	1836	111
8	406	1201	1548	149
9	279	997	1295	170
10	205	734	989	233
11	100	612	811	228
12	73	438	591	262
13	32	328	439	300
14	32	235	334	273
15	15	181	243	276
16	17	123	169	249
17	8	97	125	229
18	1	47	82	239
19	3	42	53	205
20-24	2	51	78	687
25-29	4	2	11	316
$\geq$30	0	1	1	220

Appendix Table 3: Example of expanded dataset and artificial censoring for emulation of Target Trial #2 (arms $x = 1$ and $x = 2$ shown; $K = 3$). Gray shading indicates intervals prior to eligibility assessment.

Surgeon ID.clone	Arm	Interval	Operative volume	Change from pre-baseline volume (x)	Mortality proportion	Artificial censoring
#1.1	-	-2	3	-	1/3	0
#1.1	-	-1	1	-	0/1	0
#1.1	$x = 1$	0	2	1	1/2	0
#1.1	$x = 1$	1	2	1	0/2	0
#1.1	$x = 1$	2	2	1	0/2	0
#1.1	$x = 1$	3	2	1	0/2	0
#1.1	$(x = 1)$	4	6	5	2/6	0
#1.2	-	-2	3	-	1/3	0
#1.2	-	-1	1	-	0/1	0
#1.2	$x = 2$	0	2	1	1/2	0
#1.2	$x = 2$	1	2	1	0/2	1
#3.1	-	-2	5	-	1/5	0
#3.1	-	-1	2	-	0/2	0
#3.1	$x = 1$	0	3	1	1/3	0
#3.1	$x = 1$	1	0	-2	0	0
#3.1	$x = 1$	2	3	1	0/3	0
#3.1	$x = 1$	3	1	-1	0/1	1
#3.2	-	-2	5	-	1/5	0
#3.2	-	-1	2	-	0/2	0
#3.2	$x = 2$	0	3	1	1/3	0
#3.2	$x = 2$	1	0	1	0	1

Appendix Table 4: Positivity check for Target Trial #1

Volume category	History: 0 to 4	History: 5 to 9	History: 10 to 14	History: 15+
-5	0	647	510	352
-4	174	889	533	311
-3	521	1288	587	318
-2	1104	1496	611	284
-1	1750	1493	592	238
0	1736	1486	521	246
1	1689	1327	499	193
2	1298	1051	383	176
3	982	867	309	157
4	695	616	276	118
5	456	454	193	93

Appendix Table 5: Positivity check for Target Trial #2

Volume category	History: 0 to 4	History: 5 to 9	History: 10 to 14	History: 15+
-5	0	161	17	6
-4	136	59	16	11
-3	199	66	18	4
-2	349	93	19	8
-1	506	89	6	4
0	176	83	15	4
1	131	44	13	3
2	102	37	10	2
3	38	25	3	1
4	24	11	1	4
5	21	8	2	0

Appendix Table 6: Mortality estimates for sensitivity analysis of Target Trial #2 (IPW)

Change in volume	90-day mortality (%)
-5	6.9 (5.6-7.0)
-4	6.8 (5.7-6.6)
-3	6.7 (5.8-6.5)
-2	6.6 (5.7-6.5)
-1	6.5 (5.7-6.4)
0	6.3 (5.6-6.2)
1	6.0 (5.5-6.1)
2	5.7 (5.4-6.0)
3	5.3 (5.2-5.9)
4	4.9 (5.0-5.8)
5	4.5 (4.8-5.9)

Appendix Table 7: Mortality estimates for Target Trial #2 (IPW; each arm separate)

Change in volume	90-day mortality (%)
-5	4.8 (3.4-8.9)
-4	5.8 (3.5-11.9)
-3	5.6 (4.1-13.0)
-2	6.5 (4.9-10.5)
-1	4.8 (3.7-14.5)
0	6.1 (4.5-10.5)
1	9.0 (5.2-19.4)
2	5.0 (3.9-14.0)
3	8.6 (3.8-37.7)
4	8.1 (4.6-55.3)
5	43.6 (12.7-73.1)

Appendix Table 8: Mortality estimates for Target Trial #4 (IPW; each arm separate)

Change in volume (added for <median, subtracted for $\geq$ median)	90-day mortality (%)
-5	3.5 (2.7-22.7)
-4	2.3 (2.3-18.4)
-3	4.1 (3.5-25.3)
-2	5.5 (4.5-16.7)
-1	8.0 (5.3-15.3)
0	6.1 (4.4-10.5)
1	4.8 (3.4-17.1)
2	7.2 (4.4-12.9)
3	6.8 (4.5-17.9)
4	9.3 (6.0-20.1)
5	5.9 (5.0-40.5)

9 Funding

This work was funded by the National Institute on Aging of the National Institutes of Health [grant number F32 AG064831-01 to A.L.M.].

10 Data Availability Statement

Data used in this study are available from the Centers for Medicare and Medicaid Services. Restrictions apply to the availability of these data and they are not publicly available. However, may be made available from the Centers for Medicare and Medicaid Services.

11 Acknowledgments

None

12 References

1. Ross JS, Normand SLT, Wang Y, et al. Hospital volume and 30-day mortality for three common medical conditions. The New England Journal of Medicine 2010;362:1110–8.
2. Haneuse S, Buist DSM, Miglioretti DL, et al. Mammographic interpretive volume and diagnostic mammogram interpretation performance in community practice. Radiology 2012;262:69–79.
3. Birkmeyer JD, Stukel TA, Siewers AE, Goodney PP, Wennberg DE, and Lucas FL. Surgeon volume and operative mortality in the United States. The New England Journal of Medicine 2003;349:2117–27.
4. Sternberg S. Hospitals Move to Limit Low-Volume Surgeries. US News & World Report. 2015. URL: https://www.usnews.com/news/articles/2015/05/19/hospitals-move-to-limit-low-volume-surgeries (visited on 01/19/2020).
5. Glance LG, Dick AW, Osler TM, and Mukamel DB. The relation between surgeon volume and outcome following off-pump vs on-pump coronary artery bypass graft surgery. Chest 2005;128:829–37.
6. Ch'ng SL, Cochrane AD, Wolfe R, Reid C, Smith CI, and Smith JA. Procedure-specific Cardiac Surgeon Volume associated with Patient outcome following Valve Surgery, but not Isolated CABG Surgery. Heart, Lung and Circulation 2015;24:583–9.
7. Hernán MA and Robins JM. Per-Protocol Analyses of Pragmatic Trials. The New England Journal of Medicine 2017;377:1391–8.
8. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. Mathematical Modelling 1986;7:1393–512.
9. Hernán MA, Hernández-Díaz S, and Robins JM. A Structural Approach to Selection Bias: Epidemiology 2004;15:615–25.
10. Hernán MA and Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. American Journal of Epidemiology 2016;183:758–64.
11. HCUP National Inpatient Sample (NIS). Healthcare Cost and Utilization Project (HCUP). Agency for Healthcare Research and Quality, Rockville, MD., 2014. URL: www.hcup-us.ahrq.gov/nisoverview.jsp.
12. Tsugawa Y, Jena AB, Orav EJ, et al. Age and sex of surgeons and mortality of older surgical patients: observational study. BMJ (Clinical research ed.) 2018;361:k1343.
13. Warehouse CCC. Chronic Conditions Data Warehouse: Condition Categories. URL: https://www2.ccwdata.org/web/guest/condition-categories (visited on 05/15/2020).
14. Cain LE, Robins JM, Lanoy E, Logan R, Costagliola D, and Hernán MA. When to start treatment? A systematic approach to the comparison of dynamic regimes using observational data. The International Journal of Biostatistics 2010;6:Article 18.
15. Orellana L, Rotnitzky A, and Robins JM. Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, Part I: main content. The International Journal of Biostatistics 2010;6:Article 8.