Operative Volume and Mortality: Patient Selection of Surgeon and Hospital

Main text, word count: 4,499
Abstract, word count: 263
Acknowledgments: none

ABSTRACT

Objective: The effect of surgeons' and hospitals' operative volumes on post-operative patient outcomes has been studied for decades and holds important policy implications. However, in many analyses of a general "volume-outcome relationship," no specific intervention is clearly defined. Without a well-defined comparison, effect estimates may be not be meaningful. Misinterpretations of such biased results may lead to unintended consequences for resulting policy interventions or patient recommendations. Our objective was to describe an approach to evaluate the effect on post-operative mortality of interventions on patients that specify a range of different surgeon and hospital volumes.

Data source: Fee-for-service Medicare claims (100%), 2011-2016.

Methods: For Medicare patients undergoing pancreatectomy, we specify four different target trials (i.e., hypothetical randomized trials) to estimate the effect on 90-day mortality of assigning patients to surgeons and hospitals with different operative volumes. Previous observational analyses can be viewed as an attempt to emulate our first two target trials.

Principal findings: The emulation of target trials with unrealistic, ill-defined interventions (e.g., those that ignore travel time) shows an apparent decrease in mortality with increasing operative volume. However, once the target trials are specified using interventions that could be possibly implemented in the real world, the effect estimate for operative volume becomes closer to null. We clarify that standard observational analyses are equivalent to the emulation of unrealistic target trials whose results are difficult to interpret and prone to bias.

Conclusions: When studying health systems interventions with observational data, the target trial framework may be of value to health services researchers to articulate causal questions that correspond to well-defined interventions for the real world.

***K*eywords** Operative volume · Hospital, High-Volume · Hospital, Low-Volume · Post-operative period · Target trial · Pancreatectomy · Health services

What is known on this topic:

- Previous analyses identified a correlation between operative volume and patient mortality
- However, previous studies were cross-sectional, did not specify well-defined interventions, and did not consider positivity violations due to travel time
- As such, for decision-making purposes, the relevance of previous estimates of the relationship between operative volume and patient outcomes is questionable

What this study adds:

- We evaluated the effects of interventions that specified a range of different surgeon and hospital volumes through the emulation of target trials
- Previous analyses can be seen as cross-sectional attempts to emulate trials with poorly-defined interventions
- The emulation of realistic trials yielded modest point estimates of the effect on mortality of surgeon and hospital volume

1 Introduction

Efforts to make surgery safer often focus on operative "volume," or the frequency with which a surgeon or hospital has performed a certain operation during, say, the previous year. Understanding the impact of operative volume on patient outcomes would be of interest to multiple parties. Health systems may seek to reduce complications by regionalizing certain operations and limiting credentialing privileges for those operations to hospitals and providers, respectively, who have exceeded minimum volume standards. Patients may consider traveling great distances to reach a higher volume surgeon or hospital, with the thought that this added experience may confer a lower operative risk. For payers, higher volume providers may represent a more sensible return on investment.

Previous studies have suggested that higher hospital and surgeon volumes have a beneficial effect on post-operative outcomes. A correlation between the "number of surgical procedures done in a hospital and the in-hospital mortality rate for those patients" was first reported in 1979.[1] Subsequently, a number other studies[2–4] echoed these findings, including a large-scale study of Medicare beneficiaries concluding that:

> "...patients can often improve their chances of survival substantially, even at high-volume hospitals, by selecting surgeons who perform the operations frequently."[5]

In response to these reports, patient safety organizations like the Leapfrog Group recommended minimum hospital and surgeon volume standards for credentialing certain procedures.[6,7] For example, these guidelines stipulated that hospitals (and surgeons) should perform at least 20 (and 10) pancreatic resections for cancer in a given year in order to safely offer that operation to patients.[7] Similar guidelines were specified for other operations. These recommendations have since been adopted by several major academic hospital networks, which have instituted "minimum-volume standards that bar hospitals in their systems from performing certain procedures unless both the hospitals and their surgeons do them often enough to keep their skill level up."[8]

The goal with such policies is to improve post-operative outcomes by restricting surgeons and hospitals from performing procedures that they infrequently perform. However, the findings that support these rules are based on relatively simplistic comparisons of high- vs. low-volume providers and institutions. As a consequence, these results cannot be measured against the number of unintended sequelae that may occur if similar policies were more widely adopted.[9] For example, regionalization may impede access-to-care for patients who would face significant travel away from home and family support to reach a "Center of Excellence" or "high-volume" surgeon. Many patients would have to increase travel times to reach higher volume hospitals,[10,11] despite evidence that a substantial proportion of patients place a strong utility on receiving care locally, compared with traveling to higher volume regional centers.[12] Here we describe how to design studies that provide actionable information about the impact of volume on patients' outcomes. To accomplish this goal, these studies need to compare policies that take into account travel times. We consider the example of patients planning to undergo elective resection of pancreatic malignancy. We estimate the effect on mortality of patients selecting a surgeon and/or hospital with different operative volumes in four hypothetical randomized trials ("target trials") of increasing complexity. We then emulate each of these target trials using observational data. In sequentially considering these four scenarios, we demonstrate how the target trial framework may be of value to health services researchers to avoid bias, inform variable selection, and guide interpretation.

2 Specification of the Target Trials

This section outlines the design of four hypothetical target trials in which individuals first diagnosed with pancreatic malignancy are randomly assigned to select a surgeon with a particular operative volume in the past year. Individuals will be eligible if they are over age 65, do not have evidence of metastatic disease, and are not expected to receive chemotherapy pre-operatively. We specify 26 groups defined by surgeon operative volumes 0, 1, 2, ..., 25. Some trials subsequently randomly assign participants to select a hospital with a particular operative volume in the past year, corresponding to one of 46 intervention arms, defined by hospital operative volumes 0, 1, 2, ..., 45. All patients will be followed for three months after assignment and the risk of all-cause mortality will be compared between groups.

Each of the four trials will assign patients to a surgeon and hospital in a different way, as described below. For all trials, we assume that we will have knowledge of the number of pancreatic resections for malignancy that each surgeon and each hospital in the United States had performed over the past year on any given date. Similar statistics have been made available by organizations such as ProPublica[13] and the New York Department of Health[14]. We then additionally ask trial participants which surgeon they would select, were we to assign them to a particular operative volume and instruct them to select a surgeon who performed exactly that number of operations in the past year. For some of the trials, we make the same request of participants regarding which hospital they would select.

Note that the surgeon may deem some patients inoperable due to peri-operative risk from comorbidities (e.g., severe congestive heart failure), short life expectancy, or extent of malignancy (i.e., surgeon judgment as to whether a tumor is borderline resectable versus unresectable). For those who do undergo surgery, the period between diagnosis and tumor resection for those who undergo surgery may vary. The time of operation may depend on the surgeon's schedule and the hospital's available operating room time. In the interim prior to initiating the pancreatic resection, some patients may develop progression to unresectable disease or severe comorbidities and the operation may be canceled as a result.

The first two target trials outlined below would yield uninterpretable results, were they in fact performed. However, these trials reveal important concepts. The next two target trials would produce more meaningful results.

2.1 Target Trial #1: Static point intervention with assignment to surgeon volume

Target Trial #1 is a 26-arm randomized trial in which eligible individuals are assigned to a surgeon with volume s, ranging from 0 to 25 pancreatic resections per year. Patients must select their surgeon from available surgeons with their assigned volume and must be ready to travel, sometimes long distances, to reach that surgeon. Provided that they adhere to the operative volume assignment, we allow patients to select their surgeon on criteria of their choosing (based on, for example, proximity, reputation, or years in practice). The causal diagram in Appendix Figure 1 represents key components of this trial.

This target trial would be of interest to neither policymakers nor patients because the interventions are poorly-defined. To see this, consider two hospitals: Hospital X performed 300 pancreatectomies in the past year, Hospital Y only performed 30 pancreatectomies in the past year. The research team at each hospital randomly assigns patients to undergo the operation by a surgeon who performed either one pancreatectomy in the past year or 10 such operations in the past year (this study may take years or decades to enroll an adequate number of patients, but that is beside the point). Despite identical protocols, estimates from the trials at Hospitals X and Y may differ greatly from one another because the operative volume of each hospital corresponds to different versions of the same intervention. For example, at Hospital X, staff may be intimately familiar with the procedure and peri-operative care and colleagues may be able to assist a less experienced surgeon when necessary.

Given the poorly defined intervention (insofar as it neglects important specifications with respect to hospital volume), the estimates from a study like Target Trial #1 are a heterogeneous blend of the effects of surgeon operative volume at hospitals with different characteristics. Target Trial #1 may not provide useful information for decision-making if hospital volume were responsible for a portion of the observed effect.

2.2 Target Trial #2: Static point intervention with assignment to surgeon volume and hospital volume

Target Trial #2 is a 1196-arm randomized trial in which eligible patients are assigned to a surgeon with volume ranging from 0 to 25 pancreatic resections per year at a hospital with volume ranging from 0 to 45 pancreatic resections per year. This trial is represented by the causal diagram in Appendix Figure 2.

Target trial #2 resolves the problem of heterogeneous interventions that vary with hospital volume, but patients may still be required to travel, sometimes long distances, to reach a surgeon and hospital fulfilling their assignment criteria. Though certain insurance companies may be willing to pay for travel to undergo an operation, regardless of distance, many participants may not be willing or able to comply with an assignment that requires long-distance travel. Therefore, implementing this trial would be impractical. A more realistic version of Target Trial #2 would take into account the distance between a patient and a surgeon/hospital with the requisite operative volumes.

2.3 Target Trial #3: Dynamic point intervention with assignment to surgeon volume and hospital volume

Target Trial #3 is a 1196-arm randomized trial in which eligible individuals are assigned to a surgeon with volume ranging from 0 to 25 pancreatic resections per year at a hospital with volume ranging from 0 to 45 pancreatic resections per year, unless the surgeon and hospital are ≥ 1.5 hours driving distance away, in which case the individual is assigned to undergo an operation by the maximal volume surgeon and hospital within 1.5 hours driving distance. If no surgeon or hospital were available within 1.5 hours, the individual is not restricted to undergo an operation by a surgeon and hospital with any particular operative volumes. Such individuals may select their surgeon and hospital by criteria of their choosing. This trial is represented by the causal diagram in Appendix Figure 3.

Target Trial #3 is a more realistic protocol to quantify the effect of assignment to a surgeon and hospital of particular volumes. The results of this trial would be useful for decision makers. However, this trial may yield results that are hard to interpret. To see this, suppose there exist certain savvy patients who know how to select the best available surgeon or hospital (among those available to them under their assigned strategy). If these savvy patients are also healthier, then

there will be an association between the specific surgeon (or hospital) selected by a patient and that patient's probability of mortality. Trial #3 does not differentiate effects due to surgeon or hospital operative volume and effects due to other characteristics of surgeons or hospitals by which individuals select them, which may be unequally distributed across volume levels and associated with particular providers. Estimates from Trial #3 would not be transportable to a future population (or another country's population) in which distribution of these measured characteristics of surgeon or hospital were to differ from the original population.

2.4 Target Trial #4: Dynamic point intervention, with assignment to surgeon volume, specific surgeon, hospital volume, and specific hospital

Target Trial #4 is a 1196-arm randomized trial in which patients are first assigned to one of the interventions specified in Target Trial #3, and then, among available surgeons and hospitals meeting the criteria, patients are again randomized to a particular surgeon and hospital. In this way, the effect of assignment to a particular operative volume of a surgeon and hospital is decoupled from the observed distribution of surgeon and hospital characteristics. (For another example of sequentially randomizing to evaluate the effects of different versions of treatment, in the setting of organ transplantation, see Wanis and colleagues, 2019.[15]) A directed acyclic graph for Target Trial #4 is shown in Appendix Figure 4.

2.5 Analysis of the target trials

Had each of the four target trials been implemented, risk of 90-day mortality under each intervention could be non-parametrically estimated:

$$\text{logit}\Big(\Pr(Y=0\big|R_{S=s,H=h}=r_{S=s,H=h})\Big)=\alpha_0+\alpha_1^T r_{S=s,H=h} \tag{1}$$

where Y is an indicator for 90-day mortality, $R_{S=s,H=h}$ denotes a standard basis indicating assignment to the joint intervention arm of a surgeon with volume s (for all trials) and a hospital with volume h (for trials 2-4). The intention-to-treat analysis compares the estimated risks under each strategy $r_{S=s,H=h} \in R_{S=s,H=h}$.

Alternatively, given the large number of intervention arms (and especially if only a moderate number of participants are enrolled), we may wish to make parametric assumptions that smooth across intervention arms and reduce the variance of our estimand. One way to accomplish this would be to fit the following logistic model:

$$\text{logit}\Big(\Pr(Y=0\big|R_S=r_S,R_H=r_H)\Big)=\alpha_0+\alpha_1^T f(r_S)+\alpha_2^T f(r_H)+\alpha_3 r_S r_H \tag{2}$$

where Y is an indicator for 90-day mortality, R_S denotes surgeon volume (for all trials), R_H denotes hospital volume (for trials 2-4), and $f(\cdot)$ denotes a flexible functional form using restricted cubic splines with 4 knots.

3 Emulation of the Target Trials

The above target trials may never be conducted because of the difficulty of having many thousands of patients adhere to so many interventions. Hence we can attempt to emulate these target trials using observational data from Medicare, a federal health insurance program which provides coverage for approximately 96% of United States citizens age 65 or older.[16]

Operative information, specifically whether beneficiaries underwent pancreatectomy for the indication of pancreatic malignancy, was obtained from the Medicare Inpatient Claims file based on International Classification of Diseases diagnosis and procedure codes. The surgeon who performed each procedure was identified via a unique provider identification number designated by the "primary operator" field of the inpatient Medicare data.[17] We used data from the American Hospital Association Annual Survey of Hospitals to determine the characteristics of hospitals.

History of relevant comorbidities was obtained the Medicare Master Beneficiary Summary File.[18] Estimated travel time by driving was obtained using OpenStreetMap© through the interface osrm in R.[19,20] Isochrone polygons containing all locations within 90 minutes were created around each included hospital based on ZIP code centroids. Patient ZIP code centroids were then queried to determine if the driving travel time to a hospital was within 90 minutes.

3.1 Eligibility criteria

We identified 9136 individuals who met the eligibility criteria between January 1, 2012 and September 30, 2016. We extracted information on patient demographics and comorbidities. Table 1 presents an overview of the data. Table 2 summarizes the characteristics of surgeons and hospitals.

3.2 Strategies

Eligible individuals were assigned to the strategy that was consistent with their observed data. Because we had to assign individuals to strategies based on surgeon and hospital volume at the occurrence of surgery, the emulation does not include patients who were assigned to a surgeon but did not ultimately undergo an operation. Therefore, our analyses are conducted under the assumption that there is no selection bias due to this restriction. We could not evaluate the plausibility of this assumption because, in the Medicare data, recording of the time of diagnosis was inconsistent. For example, the earliest date of diagnosis was on the day of surgery for 60% of patients, which is unrealistic. Despite this, surgeon volume was similar across categories of time from diagnosis to operation: the mean surgeon volume was 8.1, 7.7, 7.8, 8.3 and 10.6 operations for patients who had time from diagnosis to surgery of <1 week, 1 week to 1 month, 1-3 months, 3-6 months, and >6 months, respectively.

Because operations performed for individuals who were not Medicare beneficiaries are not included in the data, our measurement of operative volume may underestimate the total (Medicare and non-Medicare) operative volume of a surgeon or hospital.

3.2.1 Static strategies

In the emulation of Target Trials #1 and #2, assignment is fully defined by the operative volume of the surgeon and hospital performing each individual's operation.

3.2.2 Dynamic strategies

In the emulation of Target Trials #3 and #4, assignment is defined by both the operative volume and travel time to a surgeon and hospital meeting the specified operative volume. Among eligible Medicare beneficiaries, 15% of patients were not within 1.5 hours of *any* surgeon or hospital and thus were assigned to the surgeon and hospital volume they selected in the observed data (the natural value of the intervention). For an intervention set at a hospital volume of 5, 10, 25, 50, or 65 operations in the past year, there were 31%, 44%, 77%, 89%, and 94% of patients, respectively, who were not within 1.5 hours of a hospital with the specified operative volume. These individuals were assigned to the value of the maximum volume hospital within 1.5 hours. Likewise, for an intervention set at a surgeon volume of 5, 10, 15, or 25 operations in the past year, there were 35%, 59%, 75%, and 89% of patients, respectively, who were not within 1.5 hours of a surgeon with the specified operative volume. These individuals were assigned to the value of the maximum volume surgeon within 1.5 hours. The remainder of individuals (i.e., those who were within 1.5 hours of a surgeon and hospital with the specified operative volumes) were assigned to the surgeon and hospital operative volume specified by the intervention.

In sensitivity analyses, we replaced the travel time threshold of 1.5 hours by thresholds of 3 hours and 6 hours. A summary of the proportion of patients within the driving time thresholds by surgeon and hospital volume is presented in Appendix Figure 5.

3.3 Randomized assignment

To emulate randomization to surgeon/hospital volume, we adjusted for the following baseline covariates, which are expected to be associated with both post-operative mortality and the selection of surgeon (or hospital): age, gender, race, inpatient status at the time of operation, comorbidity history (including myocardial infarction, dementia, atrial fibrillation, chronic kidney injury, chronic obstructive pulmonary disease, congestive heart failure, and stroke or transient ischemic attack), and year of operation.

For simplicity of modeling, we assume that all combinations of surgeon operative volume and hospital operative volume could exist, within covariate patterns. However, combinations of strategies involving the highest volume surgeons and the lowest volume hospitals should be interpreted with caution. For example, at low volume hospitals performing 2 operations per year, the highest volume surgeon performed 19 operations in the past year.

For Target Trial #4, we also emulate randomization to a specific surgeon and hospital within levels of operative volume by adjusting for characteristics of available surgeons and hospitals. For surgeons, these covariates included age, gender, and working at more than one hospital; for hospitals, these covariates included availability of a cardiac intensive care unit, hospital type (for profit, not for profit, or government non-Federal), teaching category (major, minor, or non-teaching), and size (1-99 beds, 100-399 beds, or $\geq$400 beds).

3.4 Follow-up

The beginning of follow-up occurs at the time of pancreatectomy. The follow-up period extended to 90 days from the date of the operation. There was no administrative censoring, given that no individuals were included after September 30, 2016. There was no censoring from loss to follow-up, given the source of outcome data, as defined in the next section.

3.5 Outcome

Mortality within 90 days of the surgery was assessed by linkage with the Vital Status file, which is based on Medicare claims data from the Medicare Common Working File, information from family members, and information from the Railroad Retirement Board and the Social Security Administration.

3.6 Causal contrast

We estimated the observational analog of a per-protocol effect.

3.7 Statistical analysis

In the following sections, we describe the statistical analysis for all trials. Additional details are available in the Appendix.

3.7.1 Trials #1 and #2

The analysis for the emulation of target trials #1 and #2 are identical, with the exception that the random variable R_H (denoting hospital volume) is removed for the analysis for the emulation of target trial #1. We adjusted for the baseline covariates L_0 via standardization using logistic regression for each of the intervention arms R_H (denoting hospital volume) and R_S (denoting surgeon volume). The regression model was similar to that of Equation (2), but with the addition of baseline covariates. To obtain valid 95% confidence intervals in the presence of clustering at the levels of surgeon and hospital, a non-parametric cluster bootstrap procedure with 1000 resamples was implemented for the emulation of all trials.

3.8 Target Trial #3: Dynamic point intervention with assignment to surgeon volume and hospital volume

Target Trial #3 mirrors the analysis of Target Trial #2, except that the intervention is dynamic, as described above. In the previous trials, the concern that travel time may be associated with both surgeon/hospital volume and the outcome could not be rectified simply by adjusting for travel time, due to positivity violations. For example, if no individual residing >500 miles away from a surgeon with a certain operative volume underwent pancreatectomy by a surgeon with this volume, then that covariate pattern would be absent from our data and we would not be able to adjust for it without strong parametric assumptions. This potential lack of positivity is not a problem for target trial #3 because travel time is explicitly incorporated into the strategies.

3.9 Target Trial #4: Dynamic point intervention, with assignment to surgeon volume, specific surgeon, hospital volume, and specific hospital

To emulate random assignment to a specific hospital and surgeon within levels of operative volume, we modified the above procedure for Target Trial #3 by standardizing across surgeon-level and hospital-level covariates, in addition to the patient covariates that were previously included.

4 Estimates from Medicare data

4.1 Estimates

Table 3 and Figure 1 present estimates of 90-day post-operative mortality for each of the four target trials. Appendix Figure 6 presents these estimates stratified by surgeon operative volume, rather than hospital operative volume.

4.1.1 Surgeon operative volume

Table 3 also presents the corresponding risk differences in 90-day post-operative mortality, relative to a surgeon who performed two pancreatectomies per year (for Trial #1) or a surgeon who performed two pancreatectomies in the past year at a hospital that performed two, 12, or 43 pancreatectomies in the past year (for Trials #2-4). Sensitivity analysis results are summarised in Appendix Figures 7 and 8.

For Target Trial #1, the estimated risk difference in 90-day mortality was -1.2 percentage points for an operation by surgeons who performed two pancreatectomies (10^{th} percentile) in the past year (without changing the distribution of available surgeons with these volumes) compared with an operation by surgeons who performed 18 pancreatectomies (90^{th} percentile) in the past year. For Target Trial #2, the corresponding risk difference was estimated to be -1.7 percentage points at lower volume hospitals, -1.1 at middle volume hospitals, and -1.1 at higher volume hospitals. For Target Trial #3, the corresponding risk difference was estimated to be -0.6 percentage points at lower volume hospitals, -0.4 at middle volume hospitals, and -0.4 at higher volume hospitals. For Target Trial #4, the corresponding risk difference was estimated to be -0.8 percentage points at lower volume hospitals, -0.6 at middle volume hospitals, and -0.6 at higher volume hospitals.

For Target Trial #1, the estimated risk difference in 90-day mortality was -2.6 percentage points for an operation by surgeons who performed two pancreatectomies (10^{th} percentile) in the past year (without changing the distribution of available surgeons with these volumes) compared with an operation by surgeons who performed 25 pancreatectomies (95^{th} percentile) in the past year. For Target Trial #2, the corresponding risk difference was estimated to be -2.3 percentage points at middle volume hospitals, and -2.3 at higher volume hospitals. For Target Trial #3, the corresponding risk difference was estimated to be -0.5 percentage points at middle volume hospitals, and -0.6 at higher volume hospitals. For Target Trial #4, the corresponding risk difference was estimated to be -0.8 percentage points at middle volume hospitals, and -0.8 at higher volume hospitals.

4.1.2 Hospital operative volume

For Target Trial #2, there was a relatively large decrease in mortality gained by patients selecting a hospital with median operative volume (12/year), compared with a hospital with lower operative volume (10^{th} percentile: 2/year), with point estimates ranging from to -0.1 to -1.3 percentage points (depending on surgeon volume). For Target Trial #3, the corresponding risk differences between middle volume and low volume hospitals ranged from -0.1 to -0.3. For Target Trial #4, the corresponding risk differences between middle volume and low volume hospitals ranged from -0.1 to -0.3. In all trials, there was a minimal effect comparing the selection of higher volume hospitals (90^{th} percentile: 43/year) with median operative volume hospitals (12/year).

5 Discussion

We evaluated the effects of interventions that specified a range of different surgeon and hospital operative volumes. To do so, we emulated four (hypothetical) target trials.

Target Trial #1 was an unrealistic study design that specified an intervention on surgeon volume only. The large estimated effect size is therefore difficult to interpret because it disregards the practical implications of travel time and depends on unspecified characteristics of the hospital (including hospital operative volume) at which the surgeon would be operating.

The emulation of Target Trial #2, which specified a joint intervention on surgeon volume and hospital volume, yielded similar magnitudes of decreases in mortality with higher surgeon volumes. Additionally, there was a decrease in mortality conferred by avoiding the lowest volume hospitals. However, this trial still disregarded travel time and therefore yielded results that are unrealistic.

Trials #3 and #4 specified more realistic interventions that explicitly incorporated travel time. In Trial #3, the difference in mortality across surgeon operative volumes was modest, compared with Trials #1 and #2. Likewise, there was an attenuated effect of hospital volume on mortality. The intervention in Trial #3 did not specify which particular surgeon and hospital for patients to select, within levels of operative volume assignment. Therefore, the estimates do not distinguish between the effect of the surgeon's (or hospital's) volume and the effect of confounders on the relationship between a particular surgeon and 90-day mortality, which may be differentially distributed among surgeon operative volumes. But this is, in fact, inconsequential for a patient selecting a surgeon based on operative volume. Conversely, for policy-makers who must consider the mechanism behind operative volume thresholds for credentialing privileges, these estimates may be less meaningful.

Trial #4 yielded similarly modest point estimates of the effect on mortality of surgeon volume and hospital volume. Because Trial #4 accounted for surgeon and hospital characteristics (unlike Trial #3), these estimates correspond to the effect of a real-world policy on surgeon and hospital volume. The results from Target Trial #4 may be most useful for future decision-makers concerned with mechanisms behind the effects of minimum operative volume thresholds as well as the effects in populations with different characteristics.

Previous analyses of observational data can be viewed as attempts to emulate trials with a design similar to Target Trials #1 or #2.[3,5,21] Therefore, the relevance of their estimates for decision-making is questionable. In addition, most prior studies have been cross-sectional; the lack of longitudinal data makes it difficult to provide the estimates with a meaningful interpretation.

Further, even though previously reported associations between surgeon volume and post-operative mortality have been interpreted as evidence that surgeons should perform procedures "often enough to keep their skill level up," these interpretations are not warranted.[22] Neither previous observational analyses nor the ones we present here assessed the effect on patient mortality of intervening on a surgeon's (or hospital's) volume.

In summary, for decision-making in health services research, it is of critical importance to specify a well-defined intervention. We illustrated that effect estimates differ based on the target trial of interest. Introducing bias with analytic choices that deviate from the intended target trial may lead to erroneous decision-making with unintended consequences.

6 References

1. Luft HS, Bunker JP, and Enthoven AC. Should operations be regionalized? The empirical relation between surgical volume and mortality. The New England Journal of Medicine 1979;301:1364–9.
2. Bach PB, Cramer LD, Schrag D, Downey RJ, Gelfand SE, and Begg CB. The influence of hospital volume on survival after resection for lung cancer. The New England Journal of Medicine 2001;345:181–8.
3. Birkmeyer JD, Siewers AE, Finlayson EVA, et al. Hospital volume and surgical mortality in the United States. The New England Journal of Medicine 2002;346:1128–37.
4. Reames BN, Ghaferi AA, Birkmeyer JD, and Dimick JB. Hospital volume and operative mortality in the modern era. Annals of Surgery 2014;260:244–51.
5. 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.
6. Group L. Surgical Volume. 2019. URL: http://www.leapfroggroup.org/ratings-reports/surgical-volume (visited on 10/06/2019).
7. Me L. Leapfrog Hopsital Survey Bibliography: Hospital & Surgeon Volume:2.
8. Modrall JG, Minter RM, Minhajuddin A, et al. The Surgeon Volume-outcome Relationship: Not Yet Ready for Policy. Annals of Surgery 2018;267:863–7.
9. Schwartz DM, Fong ZV, Warshaw AL, Zinner MJ, and Chang DC. The Hidden Consequences of the Volume Pledge: "No Patient Left Behind"? Annals of Surgery 2017;265:273–4.
10. Birkmeyer JD, Siewers AE, Marth NJ, and Goodman DC. Regionalization of high-risk surgery and implications for patient travel times. JAMA 2003;290:2703–8.
11. Merrill AL, Jha AK, and Dimick JB. Clinical Effect of Surgical Volume. The New England Journal of Medicine 2016;374:1380–2.
12. Finlayson SR, Birkmeyer JD, Tosteson AN, and Nease RF. Patient preferences for location of care: implications for regionalization. Medical Care 1999;37:204–9.
13. Pierce O and Allen M. Assessing surgeon-level risk of patient harm during elective surgery for public reporting. 2015:21.
14. Committee NYSCA. Adult Cardiac Surgery in New York State 2014-2016:62.
15. Wanis KN, Madenci AL, Dokus MK, et al. The meaning of confounding adjustment in the presence of multiple versions of treatment: an application to organ transplantation. European Journal of Epidemiology 2019;34:225–33.
16. Research Data Assistance Center. Getting Started with CMS Data | ResDAC. URL: https://www.resdac.org/getting-started-cms-data (visited on 05/11/2020).
17. 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.
18. Warehouse CCC. Chronic Conditions Data Warehouse: Condition Categories. URL: https://www2.ccwdata.org/web/guest/condition-categories (visited on 05/15/2020).
19. contributors O. Planet dump retrieved from https://planet.osm.org. Tech. rep. 2017. URL: https://www.openstreetmap.org (visited on 02/01/2020).
20. Giraud T. osrm: Interface Between R and the OpenStreetMap-Based Routing Service OSRM. Tech. rep. R package version 3.3.2. 2019. URL: https://CRAN.R-project.org/packagae=osrm.
21. Krautz C, Nimptsch U, Weber GF, Mansky T, and Grützmann R. Effect of Hospital Volume on In-hospital Morbidity and Mortality Following Pancreatic Surgery in Germany. Annals of Surgery 2018;267:411–7.
22. Sternberg S. Hospitals Move to Limit Low-Volume Surgeries. 2015. URL: https://www.usnews.com/news/articles/2015/05/19/hospitals-move-to-limit-low-volume-surgeries (visited on 01/19/2020).

7 Tables

Table 1. Characteristics of 9136 individuals who underwent elective pancreatic resection for malignancy reimbursed by U.S. Medicare, 2012-2016

	Count (%) or Median [interquartile range]
Age, years	73.3 [69.1, 78.1]
Female	4494 (49)
Race	
Black	600 (7)
White	8130 (89)
Other	406 (4)
Acute myocardial infarction	303 (3)
Dementia	435 (5)
Atrial fibrillation	1197 (13)
Chronic kidney disease	2435 (27)
Chronic obstructive pulmonary disease	1918 (21)
Congestive heart failure	1710 (19)
Diabetes	4122 (45)
Coronary artery disease	4318 (47)
Stroke or transient ischemic attack	832 (9)
Inpatient at time of operation	576 (6)
Hospital annual volume	12.0 [5.0, 23.0]
Surgeon annual volume	5.0 [3.0, 10.0]
Year	
2012	1810 (20)
2013	1885 (21)
2014	1864 (20)
2015	1944 (21)
2016	1633 (18)

Table 2. Characteristics of surgeons and hospitals who performed elective pancreatic resection for malignancy reimburssed by U.S. Medicare, 2013-2016

	Overall
No. hospitals	697
Mean (range) annual hospital volume	6 (1-139)
Cardiac Intensive Care Unit (%)	505 (72)
Hospital type	
For profit	89 (13)
Not for profit	534 (77)
Govt. non-Federal	74 (11)
Teaching category	
Major	210 (30)
Minor	278 (40)
Non-teaching	209 (30)
Size	
Small (1-99 beds)	15 (2)
Medium (100-399 beds)	342 (49)
Large ($\geq$ 400 beds)	340 (49)
No. surgeons	1358
Mean (range) annual surgeon volume	3 (1-50)
Mean (range) surgeon age, years	50 (31-84)
Female surgeons (%)	105 (8)
Operate at more than 1 hospital (%)	226 (17)

Table 3. Estimates and risk differences (95% confidence interval) in 90-day mortality by surgeon and hospital volume for each target trial. (Reference: lowest volume surgeon wtihin each hospital volume category.)

	Surgeon: 2/year	Surgeon: 5/year	Surgeon: 18/year	Surgeon: 25/year
Hospital: Natural value				
Trial 1: Estimates	7.9 (6.4, 9.4)	7.5 (5.9, 9.5)	6.6 (4.4, 10.2)	5.2 (2.7, 10.9)
Risk differences		-0.3 (-2.1, 1.9)	-1.2 (-4.5, 3.0)	-2.6 (-5.7, 3.2)
Hospital: 2/year				
Trial 2: Estimates	11.0 (9.1, 12.9)	10.5 (7.5, 14.4)	9.3 (5.5, 14.6)	7.4 (3.7, 14.7)
Risk differences		-0.4 (-2.6, 2.5)	-1.7 (-5.7, 3.4)	-3.6 (-7.6, 3.5)
Trial 3: Estimates	10.5 (8.9, 12.4)	10.2 (7.9, 13.2)	10.0 (7.6, 13.3)	9.7 (7.5, 12.8)
Risk differences		-0.3 (-1.9, 1.7)	-0.6 (-2.5, 2.2)	-0.8 (-2.7, 1.9)
Trial 4: Estimates	10.0 (8.0, 12.3)	9.5 (7.2, 12.9)	9.1 (6.7, 12.4)	8.9 (6.6, 12.2)
Risk differences		-0.5 (-1.9, 1.6)	-0.8 (-2.5, 1.3)	-1.1 (-2.7, 1.1)
Hospital: 12/year				
Trial 2: Estimates	6.8 (4.8, 9.0)	6.5 (5.1, 8.3)	5.7 (3.4, 8.9)	4.5 (2.1, 9.3)
Risk differences		-0.3 (-2.1, 1.4)	-1.1 (-4.0, 2.2)	-2.3 (-5.1, 2.2)
Trial 3: Estimates	7.8 (6.3, 9.3)	7.5 (6.3, 8.9)	7.4 (6.1, 8.8)	7.2 (6.0, 8.7)
Risk differences		-0.2 (-1.5, 1.0)	-0.4 (-1.9, 1.2)	-0.5 (-2.0, 1.1)
Trial 4: Estimates	7.9 (6.5, 9.6)	7.5 (6.3, 9.0)	7.3 (6.0, 8.7)	7.1 (5.9, 8.6)
Risk differences		-0.4 (-1.6, 1.0)	-0.6 (-2.1, 0.8)	-0.8 (-2.3, 0.7)
Hospital: 43/year				
Trial 2: Estimates	6.8 (3.2, 10.4)	6.5 (3.1, 10.4)	5.7 (3.1, 8.3)	4.5 (2.2, 7.2)
Risk differences		-0.3 (-1.9, 1.5)	-1.1 (-4.7, 1.6)	-2.3 (-6.0, 1.4)
Trial 3: Estimates	7.8 (6.4, 9.4)	7.6 (6.2, 9.2)	7.4 (6.3, 8.7)	7.2 (6.2, 8.5)
Risk differences		-0.2 (-1.5, 1.0)	-0.4 (-2.0, 1.1)	-0.6 (-2.2, 1.0)
Trial 4: Estimates	8.0 (6.6, 9.7)	7.6 (6.4, 9.3)	7.4 (6.4, 8.6)	7.2 (6.3, 8.4)
Risk differences		-0.4 (-1.6, 1.0)	-0.6 (-2.4, 0.8)	-0.8 (-2.5, 0.7)

8 Figure legends

Figure 1. Estimates of 90-day post-operative mortality for each target trial. (The surgeon volumes displayed range from the 5^{th} to 95^{th} percentiles of observed surgeon volume.)

9 Figures

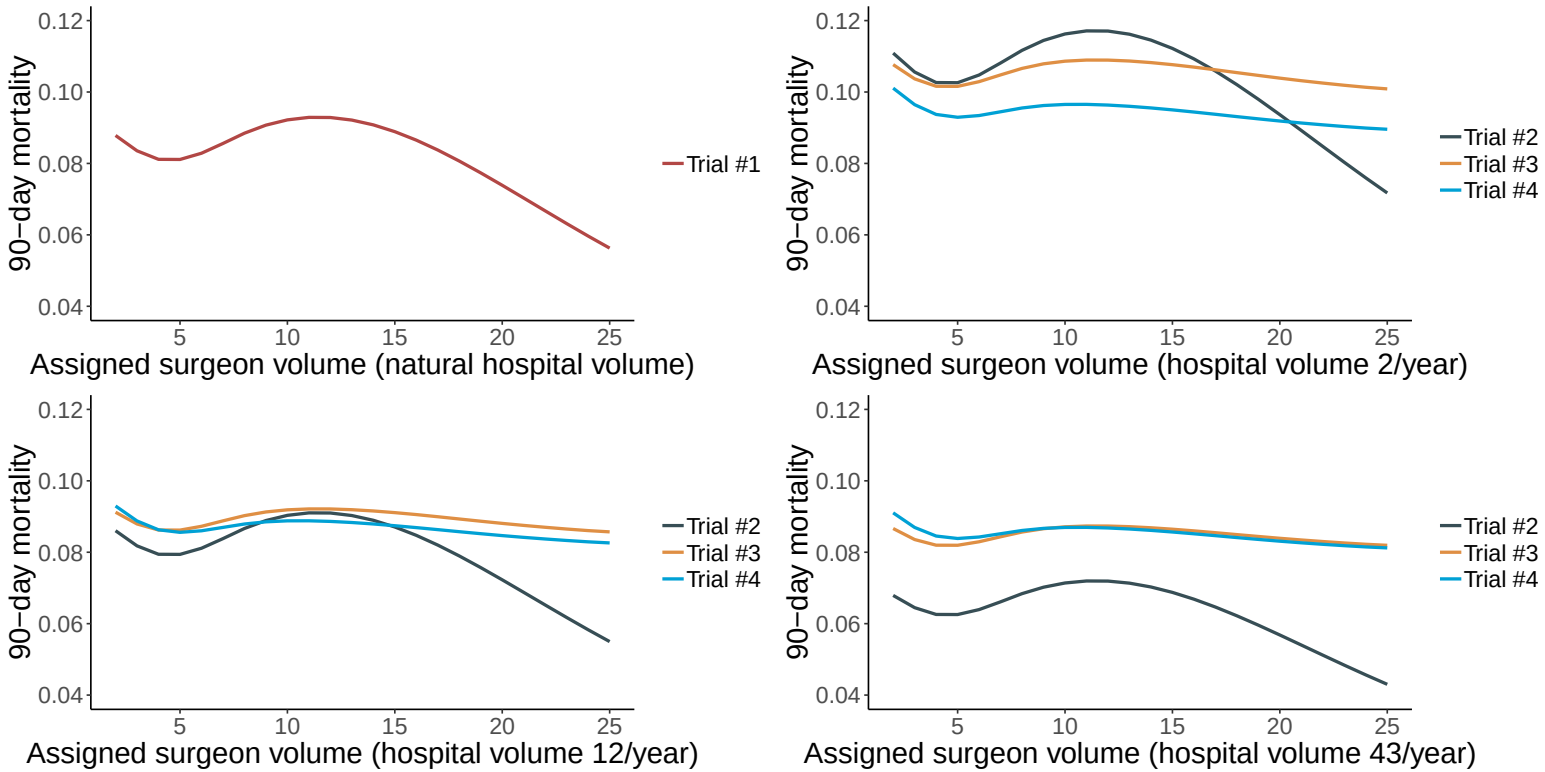
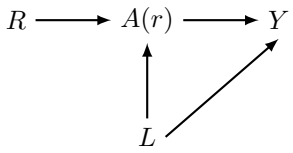


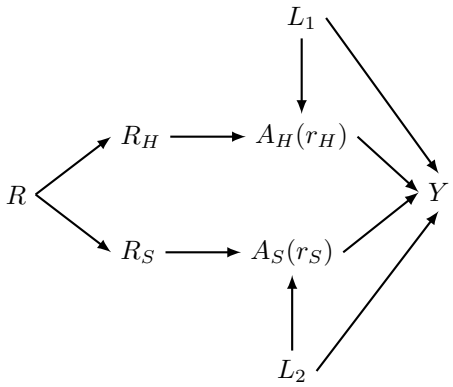
Figure 1. Estimates of 90-day post-operative mortality for each target trial. (The surgeon volumes displayed range from the 5th to 95th percentiles of observed surgeon volume.)

10 Appendix

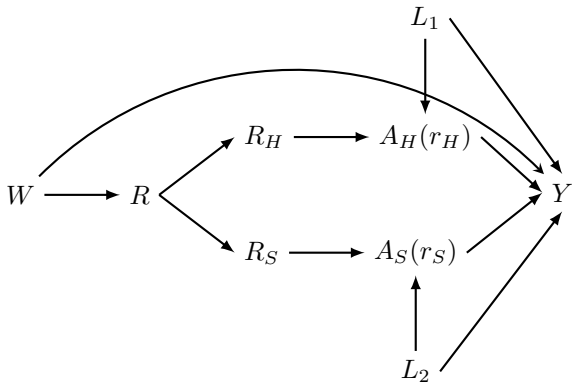
10.1 Appendix Figures



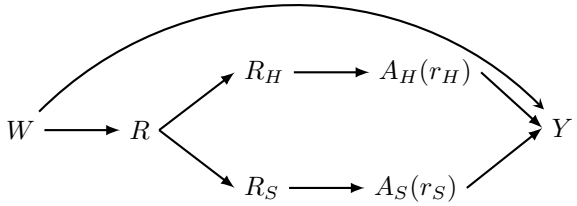
Appendix Figure 1. Directed acyclic graph for Target Trial #1, in which the compound intervention R is randomly assigned and precedes the relevant version of surgeon. Node R denotes an individual's assignment to a surgeon with a specific operative volume, $A(r)$ denotes an individual's selection of a particular surgeon with the specified operative volume r, L denotes covariates associated with both the particular surgeon selected and mortality, and Y represents mortality at the end of follow-up.



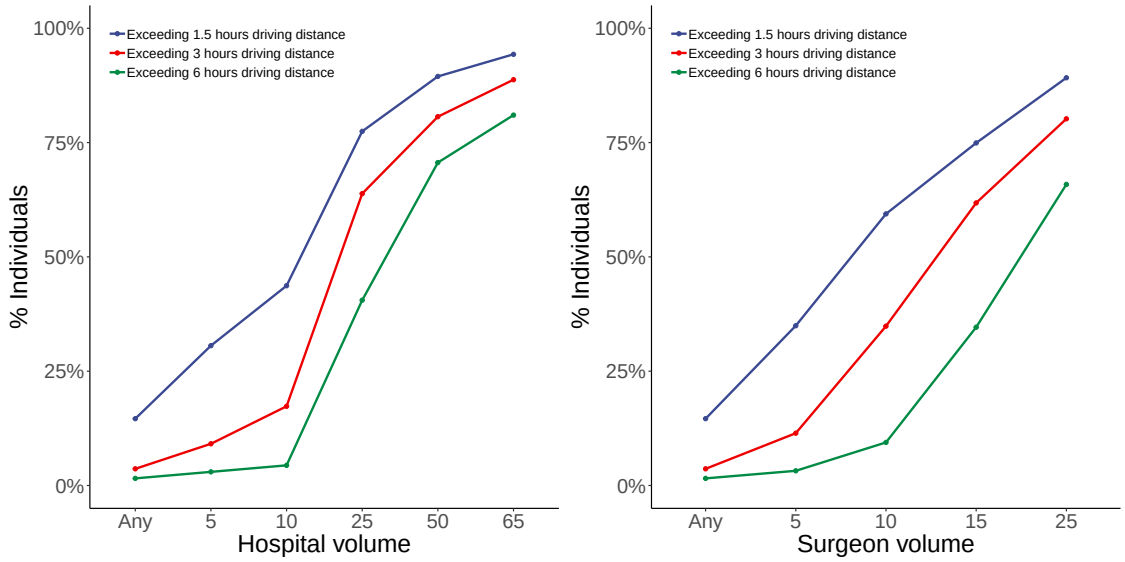
Appendix Figure 2. Directed acyclic graph for Target Trial #2, in which the compound intervention R denotes a random joint assignment of surgeon volume and hospital volume; R_H denotes the hospital volume component of R; R_S denotes the surgeon volume component of R; $A_H(r_H)$ denotes an individual's selection of a particular hospital with the specified operative volume r_H, $A_S(r_S)$ denotes an individual's selection of a particular surgeon with the specified operative volume r_S, L_1 denotes covariates associated with both the particular hospital selected and mortality, L_2 denotes covariates associated with both the particular surgeon selected and mortality, and Y denotes mortality at the end of follow-up.



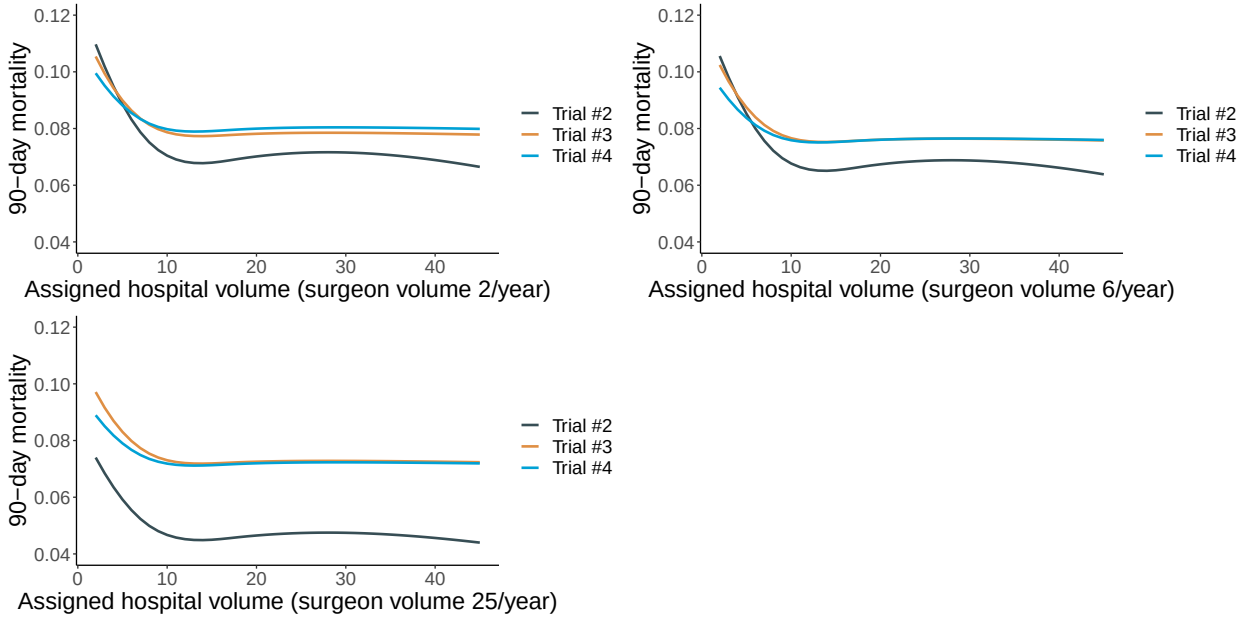
Appendix Figure 3. Directed acyclic graph for Target Trial #3, in which the compound intervention R denotes a random joint assignment of surgeon volume and hospital volume; R_H denotes the hospital volume component of R; R_S denotes the surgeon volume component of R; $A_H(r_H)$ denotes an individual's selection of a particular hospital with the specified operative volume r_H, $A_S(r_S)$ denotes an individual's selection of a particular surgeon with the specified operative volume r_S, W denotes travel time from the nearest surgeon with volume $\geq R_S$ and hospital with volume $\geq R_H$; L_1 denotes covariates associated with both the particular hospital selected and mortality, L_2 denotes covariates associated with both the particular surgeon selected and mortality, and Y denotes mortality at the end of follow-up.



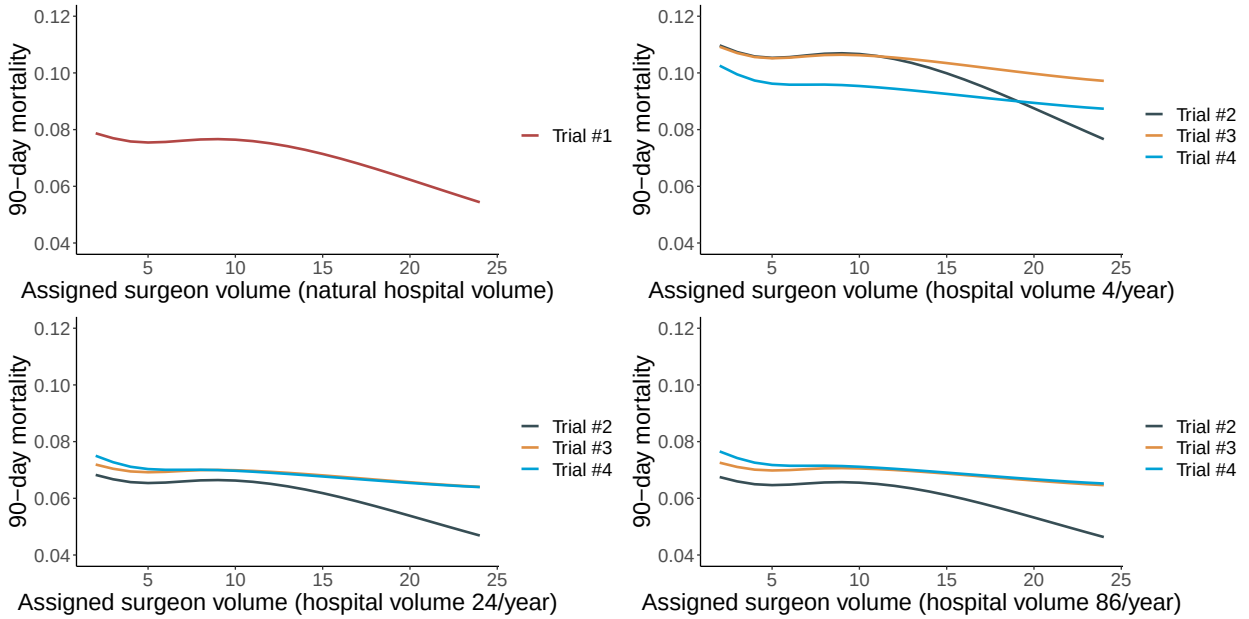
Appendix Figure 4. Directed acyclic graph for Target Trial #4, in which the compound intervention R denotes a random joint assignment of surgeon volume and hospital volume; R_H denotes the hospital volume component of R; R_S denotes the surgeon volume component of R; $A_H(r_H)$ denotes an individual's random assignment to a particular hospital within levels of the specified operative volume r_H, $A_S(r_S)$ denotes an individual's random assignment to a particular surgeon within levels of the specified operative volume r_S, W denotes travel time from the nearest surgeon with volume $\geq R_S$ and hospital with volume $\geq R_H$; and Y denotes mortality at the end of follow-up.



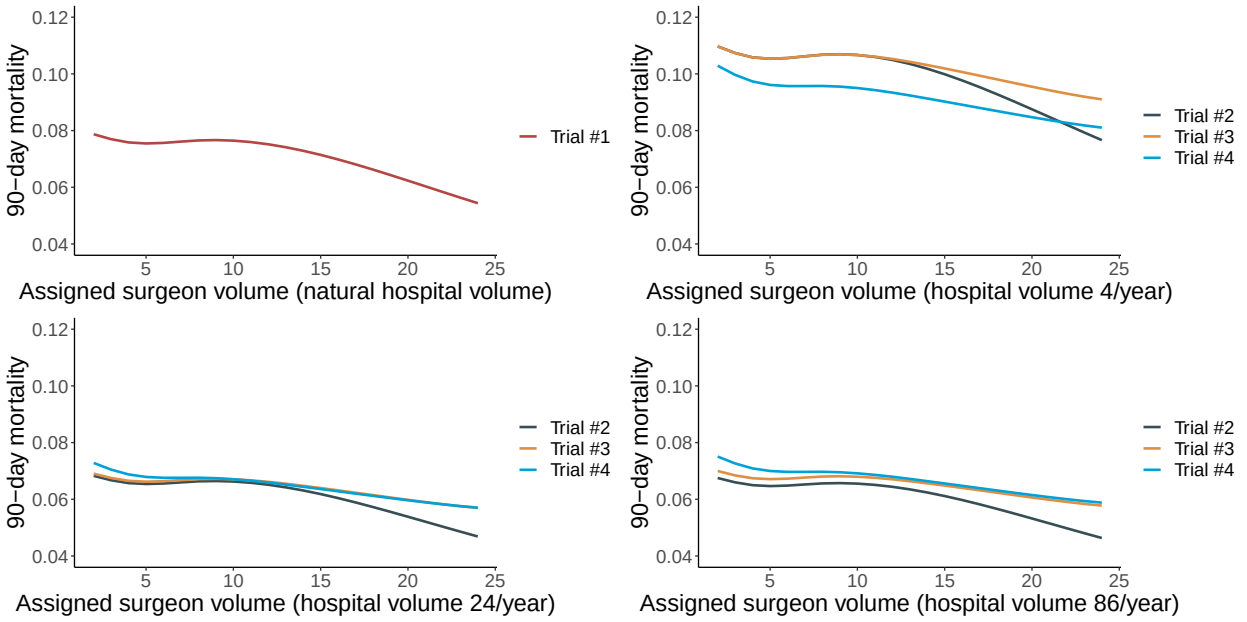
Appendix Figure 5. Proportion of individuals within the driving time thresholds by surgeon and hospital volume.)



Appendix Figure 6. Estimates of 90-day post-operative mortality for each target trial (dynamic regimes with threshold of 3 hours travel time), stratified by surgeon volume.



Appendix Figure 7. Estimates of 90-day post-operative mortality for each target trial (dynamic regimes with threshold of 3 hours travel time). (The surgeon volumes displayed range from the 5th to 95th percentiles of observed surgeon volume.)



Appendix Figure 8. Estimates of 90-day post-operative mortality for each target trial (dynamic regimes with threshold of 6 hours travel time). (The surgeon volumes displayed range from the 5th to 95th percentiles of observed surgeon volume.)

10.2 Appendix Equations

10.2.1 Trials #1 and #2

$$\sum_{l \in L} \Big[\mathbb{E}\Big(Y \big| R_S = r_S, R_H = r_H, L_0 = l_0 \Big) p(L_0 = l_0) \Big] \tag{3}$$

$$= \mathbb{E}_{l \in L} \mathbb{E}\Big[Y \big| R_S = r_S, R_H = r_H, L_0 = l_0) \Big] \tag{4}$$

One way to estimate the conditional mean in Equation (4) is by fitting the following logistic model:

$$\hat{\mathbb{E}}\Big[Y \big| R_S = r_S, R_H = r_H, L_0 = l_0) \Big] = \tag{5}$$

$$\text{expit}\Big[\alpha_0 + \alpha_1^T f(r_S) + \alpha_2^T f(r_H) + \alpha_3 r_S r_H + \alpha_4^T l_0 \Big]$$

We could then compute Equation (4) by computing the population sample average of each covariate pattern.

10.2.2 Trial #3

Given the the dynamic regime, the Equation (5) above must be modified:

$$\hat{\mathbb{E}}\Big[Y = 0 \big| R_S = g(r_S), R_H = g(r_H), L_0 = l_0) \Big] \tag{6}$$

where, for each individual, $g(r)$ takes the value of r, if the individual is within 1.5 hours of a surgeon with operative volume r; the value of the maximum volume surgeon within 1.5 hours of the individual, if the individual is >1.5 hours from a surgeon with the specified operative volume; or the observed surgeon operative volume, if the individual is not within 1.5 hours of *any* surgeon. The analogous rule is applied for hospitals.

10.2.3 Trial #4

$$\hat{\mathbb{E}}\Big[Y = 0 \big| R_S = g(r_S), R_H = g(r_H), L_0 = l_0, L_1 = l_1) \Big] \tag{7}$$

For surgeons, such covariates in L_1 included age, gender, and working at more than one hospital; for hospitals, such covariates in L_1 included availability of a cardiac intensive care unit, hospital type (for profit, not for profit, or government non-Federal), teaching category (major, minor, or non-teaching), and size (1-99 beds, 100-399 beds, or $\geq$400 beds).