Doctor's Orders: A Field Experiment on Fiscal Responsibility in Health Care*

Anna Costello[1], Henry Eyring[2], and Eric Floyd[3]

[1] *University of Chicago, Booth School of Business*
[2] *Duke University, Fuqua School of Business*
[3] *University of California San Diego*

April 2025

Preliminary and incomplete. Please do not share without authors' consent.

Abstract

To help address the rising costs of health care, we conduct a field experiment that aims to reduce the cost of supplies used in surgical procedures. Specifically, we develop an algorithm to identify plausible "swaps" in common supplies used in operating rooms, which would generate considerable savings but would not have any adverse impacts on patient outcomes (e.g., gauze). We partner with the University of Utah Health Care and Duke Health to roll out our experiment, which seeks to identify the best way to motivate doctors to agree to implement our suggested supply swaps. Our treatments are motivated by the significant attention constraints faced by O.R. doctors. Specifically, we explore whether providing salient information on cost savings options is enough to motivate doctors (i.e., information alone) or whether additional incentives are necessary to motivate cost savings behavior – our incentive treatments appeal to altruism, peer comparison, and financial incentives. We evaluate the relative effectiveness of each approach by measuring the one-click responses of doctors to accept or decline the supply swap. Should the doctors accept our proposed swaps, savings in the first year alone would be roughly $5M across the University of Utah Surgery Department and roughly $2M for the Duke Health Ambulatory Surgery Center. Overall, our paper provides causal evidence on effective strategies to align incentives in a decentralized organizational model, where agents face significant constraints in their ability to process information.

Keywords: Supply Chain; Incentives; Health Care; Limited Attention; Decentralized

*We appreciate comments from participants in John List's working group on field experiments, the Columbia Management Accounting Conference, Duke University, University of San Diego California, and the Georgetown Field Experiments in Accounting Conference. We also acknowledge feedback and from John List, Canice Prendergast, Mike Price, Dana Suskind, and senior members of the surgical teams at the University of Chicago Medicine, University of Utah Health Care, and Duke Health. We thank Blair Moore and Trent Turbyfill for their valuable research assistance.

1 Introduction

The U.S. has seen an unprecedented rise in the cost of health care over time, representing about 7 percent of GDP in 1970 and rising to 17 percent of GDP by 2023.[1] Though the surge in costs is due in large part to an aging population and an increase in chronic illnesses, a still substantive piece of the puzzle is a dramatic increase in the input costs that medical providers face in providing care. The costs of supplies, including medical and surgical instruments, medical devices, and other equipment used in hospital medical procedures increased by 26 percent from 2021 to 2024.[2] In response, hospital systems are looking for ways to curb spending, with a growing focus on physician behavior in selection of supplies used in surgery.

Physicians are given substantial autonomy in the way they perform medical procedures, ranging from the choice in the procedure and care plan, to how the procedure is executed, to which supplies are used during treatment. This autonomy is an integral part of the organizational design of most hospitals, as it is thought that decentralization puts decision rights in the hands of the most informed agents (Stein (2002); Harris and Raviv (2002); Aghion and Tirole (1997)). One can easily see the value of decentralized decision making in the context of medical care, particularly in more complex medical cases. Say, for example, a patient presents with severe chest pain and limited blood flow to the heart. It is useful for the doctor, who has expertise and gains information in her interactions with the patient, to have autonomy over which tests to run and which treatment is most suitable (e.g., inserting a stent versus conducting a coronary artery bypass graft). That is, treatment plans are not one size fits all, so decentralization is critical in the health care setting.

Autonomy, however, can also come with costs to the organization, patients, and society if doctors' incentives are not aligned across important dimensions. For instance, typically physicians and hospital management are aligned on the desire to provide high quality patient care; however, hospital management is also focused on the organization's fiscal condition whereas doctors are less concerned with the financial implications of their day-to-day activities. Since surgical procedures make up a large portion of a hospital's operating expenditures,

[1]https://www.cms.gov/data-research/statistics-trends-and-reports/national-health-expenditure-data/historical

[2]https://www.kaufmanhall.com/sites/default/files/2024-03/KaufmanHall-March2024-NHFR.pdf

doctors' choices can significantly influence hospital profitability.

Our study focuses on how to provide salient information, in combination with incentives, to encourage doctors to save costs in their operating room (O.R.) supply selections. Doing so should improve the alignment of doctors and hospital administration in fiscal decision making. Our goal is to do this without sacrificing the decentralized model, and in particular, retaining the physician's focus on high quality patient care. To do so, we implement a randomized experiment that provides nudges to doctors to select lower cost, but equivalent quality supplies. We partner with the University of Utah Health Care (UUHC) and Duke Health surgery departments to perform our experiment.[3]

As a first step in our analysis, we seek to identify and validate a viable set of alternative, lower cost supplies (henceforth supply swaps). In doing so, however, we balance the desire to keep both the doctor and the patient experience constant. That is, our supply swap suggestions should not fundamentally change the way the doctor performs the procedure, nor should the swaps adversely impact patient outcomes. Thus, we target items that are common and more seamlessly interchangeable, like bandages, syringes, scissors, etc. See, for example, Figure 1, which shows two catheters with identical specifications, one sold by Terumo for \$75 and the other sold by Boston Scientific \$40. By switching to the Boston Scientific catheter, hospitals could save 47% of the costs of catheters alone, without sacrificing the functionality of the tool and the associated patient outcomes. Though our intervention targets savings opportunities for low-cost, undifferentiated supplies, the procedures that rely on them are common and thus aggregate potential savings are substantial.

We obtain historical supply data used in procedures at UUHC and Duke Health and perform several analyses to identify swap suggestions and validate that they are appropriate. A critical aspect of our study is to ensure that our recommended supply swaps, though lower cost, do not adversely impact quality outcomes. Prior work typically fails to find that health care costs correlate with quality outcomes, though the literature is mixed (Jamalabadi, Winter, and Schreyögg (2020)). To establish whether there is a cost-quality link in the supplies, we implement statistical analyses relating several patient outcomes to the supplies that were used in the surgical procedure, controlling for patient comorbidities and procedure

[3] Authors are under agreement with University of Chicago Medicine as a third field site.

types. The majority of supplies – 98.6% – have *no measurable association* with patient outcomes in comparison to alternative supplies for the same procedure.[4] For those supplies that do have a statistical association with outcomes, we are careful to only suggest switching to *higher quality* supplies. That is, our swap suggestions reduce costs but maintain or even improve quality outcomes. As a second safeguard in our supply swap suggestions, we only suggest swaps to alternative supplies that are currently in use at the given field site by a peer doctor performing the same procedure. As a final check for our suggested swaps, we vet each suggestion with a senior member of the surgical team (e.g., the Chief of Surgery). The above analyses produce a set of supply swap suggestions at the doctor-procedure level. Our initial estimates indicate that our planned role-out of the experiment could yield a conservatively estimated savings of approximately $5M for the UUHC surgery department each year and approximately $2M for the subset of surgeons at Duke Health who operate at the Duke Health Ambulatory Surgery Center.

With our set of suggested supply swaps in hand, our experiment aims to *nudge* doctors to decide to switch to lower-cost supplies, rather than *mandating* supply switches. Doing so helps to maintain the decentralized decision model used by the organization. Providing a powerful nudge to doctors to encourage supply switches inherently begs the question of how supplies are selected absent our intervention, and why doctors do not select lower cost products to begin with. Supplies are selected for each procedure on a largely one-time, upfront basis and are specified in a form called the Doctor Preference Card (DPC). The DPC ensures that the supplies are stocked, sterilized, and ready in the O.R. before the procedure begins, and by specifying the DPC upfront this limits the time and repetitive task of selecting supplies each time the doctor enters the O.R. Anecdotally, DPCs are filled out when a doctor is hired, and are often selected based on the brands and products used in that particular doctor's residency and other training roles.

Doctors have the option of changing their DPC at any time, but changing a DPC is very rare in the data, revealing that substantial frictions exist in the current environment. When asked about the process to change a DPC, hospital staff revealed that doctors do not have

[4] Many of the supplies that have an impact on patient outcomes are implants and other technical equipment that are automatically excluded from our analyses.

direct access to supply cost data, and that they must request that a nurse leader make any requested changes. Moreover, the prices of goods fluctuate over time; something that doctors typically do not pay attention to.

Inspired by these anecdotes, our experimental interventions address the notion that decision makers face frictions related to limited attention, which is particularly true of surgeons in high volume medical centers. Surgeons face a high workload and must balance time in the O.R., time in administrative duties, and other clinical and research work (Kc and Terwiesch (2009); Berry Jaeker and Tucker (2017)). This is likely to lead to fatigue and significant processing constraints when it comes to ancillary tasks (Janhofer, Lakhiani, and Song (2019)). In other words, surgeons face high processing constraints and thus likely allocate their scarce resources to primary duties, with limited attention to tasks like supply cost savings and other administrative tasks for which they are not directly held accountable (Sims (2003); Kahneman (1973)). Our nudges thus address this limited attention constraint by both lowering the processing costs of analyzing and selecting supply alternatives and by layering on additional incentives for doctors to devote their attention to supply cost efficiency.

Our treatments seek to address the processing cost constraint faced by doctors. Our control sample receives an "information only" message that relays information about a lower cost DPC and asks whether the doctor would be willing to switch. The message is transmitted through an email sent from the Chief of Surgery, and it contains both information about the suggested swaps as well as a one-click approval/denial button that will transfer information to the administrative personnel who will update the doctor's DPC accordingly. The design of this message is intended to lower the costs associated with processing information related to supply changes and to limit the transaction costs of manually requesting these changes from a nurse leader. Because DPCs are so stable in the pre-period (e.g., changing a DPC is a rare event), a pre-post comparison of supply swaps after our intervention allows us to speak to the impact of providing information on supply selections. That is, we estimate the impact of providing salient information on supply selections by comparing the changes in DPCs for the control group that receives the information only message.

In addition to providing information about supply swaps, we propose three treatment conditions, each providing a different incentive to agree to reduce supply costs. These three treatments are guided by the literature on limited attention and seek to encourage doctors to increase their attention to cost management. First, we provide information about supply swaps and the estimated portion of the savings that will get passed on to the patient. In doing so, we test for the doctor's altruistic behavior, whereby their cost-reduction effort increases in the return to the patient (Becker (1974); DellaVigna, List, Malmendier, and Rao (2022)). In our second treatment arm, we provide information about supply swaps and a benchmark for the doctor's cost-efficiency relative to her peers. To appeal to the doctor's competitive nature, the message details how the doctor's relative cost-efficiency rank would improve should she accept the swap suggestions (Hallsworth, List, Metcalfe, and Vlaev (2017). In our third and final treatment arm, we provide information about supply swaps and a financial incentive that allows a portion of her cost savings to be transferred to her departmental budget. The third treatment approximates an incentive-pay alignment as a solution to the problem (e.g., Meckling and Jensen (1976)).

Before launching our main experiment, we ran a paid survey to elicit priors on the effectiveness of our proposed treatments. The goal of the survey was threefold. First, we wanted to pilot our proposed treatments since we would not be able to do so in our field setting. Second, the survey helps us conduct power analysis in the absence of informative pre-experimental data. Finally, the survey provides interesting empirical facts about people's beliefs about doctors and about their own behavior.

The survey was conducted on Prolific, using respondents working in the health care sector. We provide respondents with a narrative about a hypothetical hospital looking to cut supply costs but maintain quality, and we ask various questions to elicit the respondents' priors over which messaging might be most effective in convincing the decision maker to swap to lower cost supplies. Importantly, we ask the same set of questions in two blocks; one where we specify that the respondent should answer from their own perspective (i.e., "How would *you* respond to these incentives?"), and one where we ask the respondent to estimate how a representative surgeon would respond. This allows us to simultaneously test how reflections of self might differ from the expectations of others' behaviors, and it also accounts for the fact that respondents, though working in healthcare, may differ from the types of surgeons that will be the subjects in our primary field experiment.

Our survey responses justify the use of all proposed treatment arms, as respondents estimate that all three could produce incremental treatment effects of supply switching relative to the control group (i.e., the "information only" group). However, respondents' beliefs are that financial incentives will produce the strongest effects in motivating doctors to switch supplies. Overall, our pre-experimental evidence supports our proposed research design which is a single control group and three treatment arms.

The initial roll-out of our experiment began in April 2024. Response rates were swift and savings are measurable, indicating that our experiment is effectively designed and take-up is satisfactory. However, before conducting statistical tests or inferences, we will wait to hit the pre-registration parameters for sample size.

Our paper contributes to the growing focus on value-based healthcare and its related literature (e.g., Porter (2008); Teisberg, Wallace, and O'Hara (2020)). Value-based healthcare, at a broad level, focuses on improving patient outcomes while reducing extraneous costs. It includes changing the care-delivery model to focus on quality and prevention while simultaneously changing the reimbursement model.[5] For example, reimbursement models that include bundled payments are potentially more conducive to value-based care; this contrasts with the fee-for-service model, which compensates hospitals for each additional service performed (Adida, Mamani, and Nassiri (2017)). Our study highlights the importance of incentivizing doctors directly in a decentralized hospital system in which doctors are not directly responsible for supply costs.

Second, our paper contributes to the literature on processing costs and the ways in which we can address these frictions. Physicians face long hours, high workloads, and multi-dimensional tasks including teaching, research, and clinical work. They are prime candidates for attention frictions due to these time constraints. We explore a novel way to draw the doctors' limited attention to an important ancillary task: surgical supply selection and cost efficiency. We use an experimental approach to determine the best way to nudge doctors' attention to the task. In doing so, we can better speak to the causal effect of attention messaging on doctors' behaviors.

Our paper also contributes to the literature on decentralized decision making and delega-

[5] www.advisory.com

tion of authority. Hospital administrators, though they typically have formal authority over decisions, delegate most of the treatment-related decisions to physicians. Whereas most of the prior work has focused on promoting value-based medicine at the policy or reimbursement level, our study seeks to directly influence the individuals with authority over supply cost decision making – doctors (Gosnell, List, and Metcalfe (2020)). Agents in this setting have more information about the patients and there is urgency in decisions; however, some decisions are suboptimal from the point of view of the principal (Aghion and Tirole (1997)). Our paper addresses this misalignment by nudging doctors to engage in cost efficient decision making without stripping their delegated authority.

The remainder of the paper proceeds as follows. Section 2 describes our background and theory; Section 3 presents the results from a pre-experimental survey; Section 4 describes the data; Section 5 presents our experimental design; and Section 6 concludes.

2 Background and Theory

2.1 Background

Medical devices in hospitals make up a significant share of health care spending. Hospital spending as a whole is over one third of U.S. health care spending, a much larger portion than any other major category. For example, hospital spending is more than three times the spending on prescription drugs.[6] Medical supplies stand at roughly 20% of a hospital's budget, on average.[7] Spending on medical supplies in hospitals is about 7% of total U.S. health care spending; in comparison, the prescription drug market is roughly 9% of health care spending.[8]

Survey evidence on surgical supply costs suggest that much of this spending could be curtailed if doctors adopted the supply choices of their more cost-effective peers who achieve the same quality. Case studies have found that, if doctors agree to converge to a cost-effective set of supplies for a given procedure, savings range between 20% and 64% of the average supply cost for the procedure, without adverse effects on quality (Skarda, Rollins, Andrews,

[6]https://www.kff.org/private-insurance/slide/hospitals-and-physicians-represent-more-than-half-of-total-health-spending/

[7]https://www.kaufmanhall.com/sites/default/files/2024-03/KaufmanHall-March2024-NHFR.pdf

[8]https://www.kff.org/private-insurance/slide/hospitals-and-physicians-represent-more-than-half-of-total-health-spending/

McFadden, Barnhart, Meyers, and Scaife (2015); Dekonenko, Oyetunji, and Rentea (2020)). Taking these results to the aggregate, back of the envelope calculations suggest that potential savings of a supply cost-reduction effort could be roughly 3% of health care spending, 0.5% of GDP, and $136B.[9] Thus, the scope of the economic problem we are addressing is substantial.

Despite the rising costs of medical supplies, research has yet to compare methods of reducing frictions that prevent more cost-effective choices for O.R. supplies. To inform our study design, we spoke with employees at many hierarchical levels across several multi-billion dollar health care systems. We collected perspectives from chief executives, procurement managers, surgeons, O.R. nurses, and billing departments. In these conversations, we sought to understand the underlying frictions to efficient supply selection, and what incentives and/or information would be impactful in addressing those frictions.

We began by conducting a series of interviews regarding the status quo at our field site, UUHC. When a doctor joins UUHC, a nurse administrator meets with a newly hired doctor to record the desired supplies and create the DPC. The doctor typically adheres to the supplies she observed during her residency, and because there are no incentives to reduce supply costs, the doctor typically does not select alternative (e.g., more cost effective) supplies. Further, the doctor is unlikely to be familiar with alternative supplies and/or their cost structure. Thus, a path dependence arises in supply ordering patterns, originating from "attending" physicians who train medical residents, and resulting in a wide variety of DPCs for different doctors performing the same procedure.

The missed opportunity to save on these interchangeable items highlights that doctors have limited attention to allocate to the financial effects of their decisions. Consider this comment from a division chair at one academic health care system: "doctors are not going to notice a difference between these [interchangeable items]". This is striking when juxtaposed with the sustained, vast differences across doctors' orders and the wide dispersion in costs.

One reason that this behavior persists may relate to the high transaction costs of a change. For instance, there is a search cost associated with looking up alternative supplies

[9] Potential savings are estimated as follows: We take the midpoint of the savings range estimated from the case studies above (e.g., roughly 42%) and apply it to the 7% of health care spending that is on hospital medical supplies. Applying that 3% savings to total expenditures and GDP produce our back of the envelop estimates.

and their pricing, and a transaction cost to requesting a DPC update from a nurse in charge of records. Our study design allows us to test whether removing those frictions is sufficient to yield change, without the necessity of adding an incentive for doctors to change. Specifically, the most basic of our four experimental conditions provides doctors with a recommended DPC change and information on cost and quality. That condition allows them to accept with one click, significantly reducing the search and transaction costs associated with changes to the DPC.

Yet, given the scarce resources faced by doctors (e.g., time and attention), providing information about potential alternatives may not be sufficient to motivate supply savings behavior alone (Maćkowiak, Matějka, and Wiederholt (2023)). Inattention to the cost-effectiveness of one's DPCs may result from a principal-agent problem, whereby a doctor does not anticipate a great enough benefit in attending to cost savings. In other words, it may be necessary to also increase the benefits of the doctor's attention by providing incentives to switch supplies and save costs. We test three treatments that could provide motivation for a doctor to adopt a more cost-effective DPC: an appeal to altruism; peer comparison; and financial incentives.

2.2 Theory

Our first treatment condition designed to motivate a doctor is an appeal to her altruistic behavior. In addition to the information about supply options provided to the control group, those in the altruism treatment group are informed that agreeing to the cost savings results in [x% of] savings passed on to the patient. These savings are calculated by taking the average portion of the bill that is the patient's out-of-pocket cost and multiplying that by the reduction in the bill that results from the suggested cost-saving supply choice. The hospital's billing and reimbursement department validated that these estimates translate into the amount of out-of-pocket cost savings for the patient that result from a supply change.

The literature on altruism suggests that individuals derive utility from the act of giving or the provision of a public good (Becker (1974); Andreoni (1989, 1990)). This would imply that doctors, when reminded that their behavior would result in cost savings to patients, may derive enough utility from this prosocial act in order to motivate them to agree to treatment. While other studies find mixed evidence regarding appeals to altruism, in the healthcare sector this message may be particularly powerful. Doctors take an oath to uphold ethical standards and are typically "mission driven," thus leading to more altruistic tendencies.

Our second treatment condition is to present evidence to a doctor on her relative cost-efficiency rank. Specifically, for all doctors performing the same procedure, we rank them according to the total cost of their DPC. Our message will report that relative ranking, along with information about how that rank would improve should the doctor accept the supply swaps. This message incentivizes behavior both through an understanding of social norms and through competition. The literature on social norms suggests that agents may adhere to peer norms when they have an understanding of where they rank relative to a benchmark (e.g., Hallsworth et al. (2017); Ferraro and Price (2013)). Moreover, becoming a doctor is a highly competitive career path, involving selective testing and admissions processes, as well as years of post-graduate training. Providing doctors with peer comparisons may be especially effective if it appeals to the doctor's competitive spirit (e.g., Erat and Gneezy (2016); Scales Jr, Moin, Fink, Berry, Afsar-Manesh, Mangione, and Kerfoot (2016); Zhang, Zhang, and Palma (2024); Gneezy, Niederle, and Rustichini (2003)).

Our final treatment condition is to provide financial incentives to the doctor should she agree to the supply swap. Direct financial incentives have long been recognized as an efficient way to align the principal and the agent in profit maximixing behavior (e.g., Meckling and Jensen (1976). Because of legal and compliance considerations, we cannot directly share cost savings with doctors. Instead, we implement a cost-sharing agreement by depositing a portion of the cost savings from the supply switch to the doctor's departmental budget. As Gneezy, Meier, and Rey-Biel (2011) discuss, financial incentives may not work if they either crowd out intrinsic motivation, or relatedly, incentivize short term behavioral changes at the expense of long-term changes. However, because the baseline rate of DPC changes by doctors is low, we do not anticipate this as a large issue in our setting.[10]

2.3 Anecdotal support for our theoretical predictions

To motivate the study design, we sought institutional insights by talking to medical professionals outside of our eventual field-experimental sample. There was a wide range of

[10]The baseline rate of DPC changes in the period before our intervention indicates the doctors' intrinsic motivations to seek out cost-efficient supply alternatives.

opinions as to which treatment condition would be most effective. For example, the CFO of a hospital told us that he believes doctors would be persuaded by the opportunity to save money for patients if that were salient. By contrast, a surgery division chief believed that doctors would not make changes if the department did not receive some of the savings. Reflecting further divergence in beliefs, a nurse administrator remarked that doctors are very competitive and that a peer comparison would likely be most effective, and another surgeon echoed that sentiment. That surgeon pointed to a behavior she changed, after what she described as 15 years of habit, when she learned that she was an outlier in efficiency. From our conversations, it appears that there is significant divergence in beliefs on the most effective way to promote fiscal responsibility among doctors in the OR while preserving their autonomy.

3 Pre-experimental Survey

Before conducting our main experiment, we ran a paid survey to elicit priors on the effectiveness of potential treatments motivated by institutional insights. The goal of the survey was threefold. First, we wanted to pilot our proposed treatments since we would not be able to do so in our field setting. Second, we wanted to use the survey to help guide our power calculations. Finally, the survey provided interesting empirical facts about healthcare workers expectations regarding doctors' behavior and motivation.

Our survey was conducted via the Prolific online survey platform. Our sample included 286 participants that indicated they work in the health care industry. The survey consisted of two main blocks. For both blocks, participants were first made aware that a hypothetical hospital would be implementing a quality improvement initiative that consisted of providing information about supply swaps and incentives to switch to a cheaper supply with the same validated level of quality. Demographic data was collected at the end of the survey.

Participants were then asked how likely they would be to switch supplies if they were part of the hospital's quality improvement initiative for four different scenarios, mirroring our proposed control and treatments: information only (control); an appeal to altruism (treatment 1); a comparison of their cost effectiveness relative to that of their peers (treatment 2); and a message that a portion of the savings would be shared with them by allocating the funds to their departmental budge (treatment 3). The first block asked the participant to answer the questions from their own perspective (e.g., "how likely would *you* be to agree to switch supplies), while the second block asked the participant to answer the questions based on how they assumed a representative surgeon would respond. Willingness to switch was measured on a 1-7 scale. We also asked participants how certain they were of their answers, which was recorded on a 1-100 scale. The survey can be viewed in Appendix A1.

Table 1 records the descriptive statistics for the sample of survey respondents, and Figure 2 provides a map of the geographical distribution of respondents. The respondents are well represented geographically, spanning several regions of the U.S., and covering both large cities and rural locations. The average age of respondents is 36, and 79% of the respondents have at least a bachelor's degree. This suggests that the respondents are experienced and well-educated, and may be able to assess both their own likelihood of agreeing to treatment and to infer doctor's responses. Around 9% of respondents have a doctorate degree, defined as an M.D. or PhD.; thus, we are likely to be capturing at least a portion of the population that participates in procedures that we cover in our experiment.

Table 2 reports the responses from the survey questions. Several interesting results emerge from our survey. First, there is a clear pattern concerning the ranking of the effectiveness of different scenarios. Altruism, in the form of patient savings, had the highest score of 5.68. This was followed by financial incentives (5.62), peer comparison (5.26), and information only (4.84). These differences are all statistically significant at the 5% level, except altruism compared to financial incentives.

At first glance, it would appear that altruism is a highly powered type of motivation in our main experiment. However, a closer look at responses split by participant gender reveals an interesting pattern. Table 3 reports the mean responses based on gender. Women are much more likely to switch supplies under the altruism treatment, whereas men are most likely to switch supplies given financial incentives.[11] To account for this, we estimate revised scores based on the gender composition of doctors, where men are more represented (approximately 65% to 35%). When we make this adjustment, the relative ranking of scores

[11] Both groups estimate that doctors are most likely to switch supplies under the financial incentive treatment.

becomes financial incentives (5.65), altruism (5.54), peer comparison (5.06), and information only (4.64).

Another interesting pattern emerges when comparing the results between the self-reported willingness to switch and the participant's belief about how other surgeons would behave. In this scenario, the relative ranking is financial incentives (5.80), peer comparison (4.73), altruism (4.23), and information only (3.78). In free-response entries, participants confirmed that they believed doctors are resistant to change and would likely only respond if they personally benefited (e.g., one participant states 'Surgeons, despite making good money, tend to be greedy about their research or departmental funds.'). Overall, scores are lower for willingness to switch for doctors than self-reported, with the lone exception being financial incentives. Interestingly, participants appear to consider doctors as self-interested even though they are mostly employed in non-profit institutions.[12]

Overall, our survey provides several important considerations. First, gender is an important factor to account for in our experimental design. We block on gender when randomizing to ensure balance among treatments, and we will split results on gender to examine heterogeneous treatment effects. Second, no treatments are implausible and no treatments are "obvious" or guaranteed. There is a relatively consistent pattern of expectations related to the efficacy of certain treatments that we can compare to experimental results. Finally, on average elicited priors suggest that treatment effects should be stronger when incentives are present. On a scale of 1-7, where 4 is neutral, the score of 3.78 in the information only scenario suggests that doctors have a neutral likelihood of switching; a score of 5.80 when incorporating financial incentives, on the other hand, suggests that incentives will drive treatment effects.

4 Data

Our experimental design is informed by detailed analysis of historical UUHC and Duke Health data on supplies used in procedures and the outcomes of those procedures. The success of our experiment relies on the assumption that supply swap recommendations do

[12] Participants are more certain about their own willingness to switch (80%) than other doctors (70%), as expected. The exception, which confirms participant's strong priors on doctor's self-interested behavior, is financial incentives (76%).

not sacrifice patient outcomes. That is, we do not want to suggest that the doctor switch to a lower quality product, where low quality is defined as either (1) not appropriate or functional for the particular procedure, or (2) functional for the procedure but at a risk of harming patient outcomes, relative to the supplies currently in place.

To assess these criteria, we rely on historical supply data and surgical procedure data. The field sites provided us with comprehensive data of historical records of surgical tools used in procedures and outcomes of those procedures. At UUHC, these data cover the period 2014 through 2023. At Duke Surgery, we have received data from Jan 2023 to July 2024 and will extend the data pull back to a range comparable to that at UUHC. The data include variables such as which tools were used in the procedure, patient demographics, clinical characteristics, patient mortality and 30-day readmission rates.

4.1 Supply selection and patient outcomes

As a first step to understand the link between the cost of supplies and patient quality outcomes, we use the UUHC historical data to test the association between supplies used in surgery and patient outcomes such as mortality and readmission rates.

Our conceptual framework will regress patient outcomes on a dummy variable for each supply used in the procedure, while controlling for patient comorbidities (the Charlson Comorbidity Index, or CCI), the Medicare Severity Diagnosis Related Group (MS-DRG) index, the patient age, and their body mass index (BMI). We also include doctor and procedure fixed effects to absorb differences across style and procedure-types. Within this framework, supplies that have a significant (adverse) relationship with patient outcomes should not be considered for our swap suggestions. One key issue with estimating the relation between a single supply tool patient outcomes is that supplies are often used in bundles (e.g., a specific suture is always used in conjunction with a particular needle and thread), creating high multicollinearity in the data. Further, the data are of high dimensionality due to the vast number of tools used in a single surgical procedure, creating a large matrix of total possible tools for each procedure that are sparsely populated (primarily set to zero) for a large number of procedures.

To deal with these issues we rely on an Elastic Net, a machine learning algorithm that combines the features of both Lasso and Ridge Regression. It is a regularized regression technique specifically used to deal with the multicollinearity and overfitting issues in our data. The algorithm works by adding two penalty terms, L1 and L2, to the least-squares objective function. The L1 norm is used to perform feature selection, and the L2 norm is used to perform feature shrinkage.

The L1 ratio was set at 0.9, emphasizing the L1 penalty, which is instrumental in driving some coefficients to zero and thus performing explicit feature selection. This is crucial when dealing with numerous correlated predictors, as it helps in reducing the model's complexity and enhancing interpretability. The regularization strength parameter, CC, was set to 0.001 for mortality and .01 for 30-day readmission, indicating a strong regularization effect to counteract overfitting and manage multicollinearity effectively. This setting allows the model to penalize large coefficients more severely, reducing the risk of overfitting while still allowing for significant variables to emerge based on their influence on patient outcomes.

The Elastic Net's dual approach reduces the risk of model overfitting while addressing multicollinearity by enabling correlated variables to be either selected or ignored together. It aids in identifying groups of tools that collectively impact outcomes, which is vital for understanding the combined effect of tool usage. Given the presence of multicollinearity, the impact of an individual tool may not independently register as significant; instead, another tool might absorb the cumulative effect of both tools due to their correlated usage. This phenomenon is critical for informed decision-making about tool changes: if two tools are commonly used together and demonstrate a combined influence on the outcome, recommendations for tool substitution or retention would logically consider both, rather than one in isolation. This holistic approach ensures that any recommendations to adjust the surgical toolkit are based on a comprehensive understanding and evaluation of tool interactions and their collective impact on surgical outcomes, aligning with current best practices in statistical modeling and clinical decision-making.

The model's hyperparameters, including the L1 ratio and regularization strength (C), were optimized using GridSearchCV with Repeated Stratified K-Folds cross-validation. The final model, optimized for the highest Receiver operating characteristic (ROC AUC) score, indicates which specific surgical tools are statistically associated with changes in patient outcomes. Note that the L1 penalty causes tools that have no relation with patient outcomes to be dropped from the model.

Model estimation uses all historical procedures from 2014 through 2023 at UUHC, along with all supplies used in those procedures and patient characteristics, doctor fixed effects, and procedure fixed effects. We report descriptive statistics for these data in Table 4. The average BMI of a patient in our sample is 29; this is higher than the average BMI of adults in the U.S. of about 26. Mortality and 30-day readmissions are relatively rare events, representing about a 4 and 5% likelihood, respectively. There are nearly 28 unique types of tools on the average doctor preference card. We will expand this analysis to include data from Duke Health once we receive data extending back earlier than 2023. We display in Table 5 the descriptive statistics currently available at Duke Health spanning from January of 2023 through July of 2024.

Panel A of Table 6 reports the results of our logistic regression with elastic net, where the dependent variable is patient mortality. The coefficients are translated into odds ratios, and can be interpreted as follows: an odds ratio greater than 1 indicates that as the predictor increases, the event (mortality) becomes more likely. Conversely, an odds ratio less than 1 suggests that an increase in the predictor is associated with a decrease in the likelihood of the event. That is, if the odds ratio is less (more) than one the tool historically *decreases* (*increases*) the likelihood of mortality/readmission, suggesting it may be of higher (lower) quality or importance in a procedure.

The control variables behave as predicted. That is, higher patient CCI, DRG, and patient age all increase the likelihood of mortality. Of 2,036 tools used in the sample procedures, 18 tools have a statistical association with patient mortality. However, many of these items are differentiated, technical supplies like implants, endoscopies, etc. Ex ante, we expected that these items would impact patient outcomes because they play a more crucial role in the procedures; this validates our decision to focus our swap suggestions on undifferentiated, common tools. Interestingly, even some of those common items (e.g., "dressings") appear to have a statistical association with patient mortality. To qualify for our swaps, we only recommend switching *away* from supplies that have adverse effects on patient outcomes and switching *toward* supplies that have positive, or beneficial statistical associations with outcomes. In doing so, we ensure that our swaps will produce outcomes of equal or higher patient quality.[13]

Panel B of Table 6 reports the results where the dependent variable is the 30-day readmission rate. Twenty-nine tools have a significant relationship with 30-day readmission, and again many of these are critical items like implants that are excluded from our set of possible swaps.

In sum, our analyses of historical supply data indicate that *undifferentiated, common* supplies, by and large, have little to no effect on patient outcomes as measured by mortality and 30-day readmission. Our robust machine learning techniques provide comfort that the supply swaps we suggest are indeed appropriate and will not harm the procedure quality.

4.2 Stock Outs

The above analysis provides compelling correlations between supplies and patient outcomes, but the selection of supplies is an endogenous choice, and thus our analysis cannot be interpreted as having a causal effect on outcomes.

To better address this issue, we seek to identify supplies that are not endogenously selected and are plausibly exogenous to the patient outcomes. Specifically, we identify supplies that are changed on the DPC because they were stocked out. Stock outs are common, particularly in the post-COVID period when supply chains continue to face frictions. These stock-outs are out of the control of the doctor; the procurement team and nurses work to find alternative products to stock for the procedures to fill the place of the stocked out item.

If the replacement supply is less expensive (and of lower quality), it should have adverse effects on patient outcomes. In contrast, if cost and quality are not related, then the replacement supply should have little to no impact on patient outcomes.

We identify stock-outs in the data as follows. We require a supply to be in use for a given period by two or more doctors performing the same primary procedure. We then require that the supply disappear from the DPC of all doctors, for a minimum of 30 days (during which time the doctors continued to perform that same procedure). Finally, we require that the same item appears back on the DPCs at some later date (after the stock-out period). For those procedures that have a stocked out supply, we create an continuous variable, *Stockout_Dollar_Change*, which measures the change in the price of the doctor's DPC from

[13]Items marked with an asterisk are included in the suggested swaps.

the period during the stockout, relative to the period before the supply was stocked out. For procedures without stockouts, this variable is zero.

To assess the impact of exogenous stock-outs on patient quality, we estimate the following regression using OLS:

$$Y_{i,p,t,d} = \beta_1 Stockout\_Dollar\_Change_{(i,p,t,d)-(i,p,t-1,d)} + \gamma Controls_p + \alpha_i + \alpha_t + \alpha_d + \varepsilon_{i,t,d} \quad (1)$$

where Y is the quality outcome for procedure i, patient p, doctor d, and time t. In addition to our *Stockout_Dollar_Change* variable, we include patient-level controls, and fixed effects for the primary procedure, the doctor, and the quarter that the procedure was completed. Standard errors are clustered at the primary procedure level. Results from this regression are presented in Table 7. We find that a dollar value change in supplies due to stock-outs has no effect on patient outcomes, measured by mortality or 30-day readmission. The results confirm analyses from Section 4.1, and provide comfort that the supply swaps we will suggest will not adversely impact quality.

4.3 Swap Selection and Swap Descriptive Statistics

Our experiment relies on having a set of viable alternative supplies to suggest to each doctor, that are both of equal or higher quality and lower cost. We go through several steps to create a specific swap suggestion for each doctor-procedure. That is, each doctor has several potential swaps, because they perform multiple procedures and because multiple swap alternatives are viable even for the same doctor-procedure.

To create potential cost-savings swaps for each doctor's procedures, we consider the set of alternative DPCs of all doctors that perform the same procedure. If another doctor performing the same procedure has a less expensive DPC (and therefore has lower cost inputs), we consider that DPC as a viable alternative to the focal doctor's DPC. The reason for doing this is as follows: swapping tools one-for-one (e.g., one bandage for another bandage) requires researcher judgement and detailed knowledge of supply specifications and their interchangability. Doing this at scale is prohibitively costly. Instead, we consider a peer doctor's DPC as viable, as it is in use, approved, and performed for the same procedure. Thus, our suggestions are not one-for-one supply swaps, but rather involve swapping full DPC kits -

which could include adding some supplies and removing others but ultimately result in cost savings. Our filters include the following: First, we restrict the set of DPCs to those used in the same primary procedure. Second, we restrict the set of DPCs to those that are less expensive than the focal doctor's DPC. Third, We restrict the viable alternative DPCs to those that would require less than 10 tool changes (drops and/or adds), in order to limit the cognitive constraint and procedural changes on the doctor.

Our fourth step in creating viable DPCs is informed by the statistical analyses discussed in sections 4.1 and 4.2. From that analysis, it is clear that differentiated products and implants should be excluded from our set of potential swaps, since they significantly impact patient outcomes. Thus, we create a list of items that are common, undifferentiated goods. The list was selected and annotated by several doctors who have an intimate knowledge of the supply categories and their functionality. The final list of viable categories is listed in Appendix A2.[14] We require that the tools that we recommend swapping are classified as common. As a fifth and final step, we require that the suggested DPC swap gain at least $1,000 in annual savings, to make sure the swap is worthwhile to justify the transaction costs.

The resulting set of DPCs, and their associated savings (from one suggested DPC change) are described in Table 8 for UUHC, and Table 8 for Duke Ambulatory Surgery Center. At UUHC (Duke), the average savings from a DPC change for one procedure-event is $1,854 ($667). Taking the average number of times that the doctor performed that procedure at UUHC (Duke) in 2023 produces an average estimate of $49,494 ($26,834) in annual savings for a single doctor. Note that the data from Duke are for the ambulatory surgery center and only for 14 commonly performed procedures. Our plan with Duke Surgery is to begin with that smaller focus and expand. Thus, larger savings comparable to UUHC are plausible as the scope of procedures and surgeons widens.

The average number of tool changes recommended is about 5. Note that total tool changes does not equal unique tools added plus unique tools removed because total tool changes also accounts for changes in quantities of tools. Overall, potential savings are substantial.

[14] Some categories of non-differentiated supplies had statistical significance in Table 6. To qualify for a swap, a supply must be *either* not statistically associated with outcomes, *or* associated with outcomes in a positive way. That ensures that our swaps are of equal or higher quality.

Further, we anticipate sending at least eight iterations in total to each doctor with a kit that meets the above qualification. Doing so will multiply the potential savings of the intervention.

5 Experimental Design

5.1 Experimental Site

Our first field site to test these treatments is UUHC, an academic medical system that has five hospitals, 12 community clinics, and several specialty centers. The system receives over one million outpatient visits and 30,000 in-patient visits annually. The more than 1,000 physicians employed at UUHC span 20 specialties, including the most advanced neurosurgery and cancer care.

The department of surgery has 96 practicing providers and has agreed to conduct the study. The orthopedics department operates in a different building and has an additional 48 practicing providers. Leadership there has solicited a proposal to apply the study there as well. We are under agreement with the University of Chicago Medicine to conduct the study at that site. Our analyses at this stage focus on the UUHC department of surgery, from which we have received both the necessary data and approvals.

To vet the suggested swaps, we have met repeatedly with senior leadership at UUHC to review each proposed swap. Our final suggested swaps, based on the methodology presented in Section 4.3, were reviewed by a senior member of the department of surgery. All suggested swaps were approved as viable (without sacrificing quality). We have carried out that same process at Duke Ambulatory Surgery Center and have reached the phase of presenting suggested DPC changes to surgeons.

5.2 Participant Population

Our analyses represent the results from the field experiment registered under the AEA RCT Registry (Costello et al., 2024). Our sample consists of doctors from the UUHC health system. Specifically, participant doctors represent the surgery department. Doctors were screened to confirm that they performed a surgery within the past 12 months and had a DPC available to propose swaps.

5.3 Power Analysis

To ensure that we have sufficient sample size, we conduct a power analysis. We make use of our pre-rollout survey to provide parameters. Specifically, the analysis requires expectations for magnitudes and standard deviations in surgeons' willingness to change in the presence of each of our treatments.

From our pre-rollout survey data, we compare average points along the Likert scale for each treatment to identify expected sample means in different experiment conditions. Our sample mean is anchored at 4 on the Likert scale, where a surgeon is neither likely nor unlikely to accept a swap, which we place as an estimated mean of 0.5 surgeons accepting the swap. For a Likert scale mean of, say 4.8, or 20% higher than 4, we increase the expected treatment group mean from 0.5 upward by a factor of 20%, to 0.6. We then compare expected means among groups. The second moment in our analysis is the standard deviation in expectations of the willingness to accept a swap.

There are six pairwise comparisons possible for four experimental conditions. Survey respondents expected financial incentives to have a much higher mean than any alternative for surgeons, but did not expect a similar effect for themselves. Thus, for the three pairwise comparisons that involve financial incentives, the expected magnitude of differences in means are significantly larger in the scenario where respondents are predicting surgeon behavior, relative to when they are predicting their own behavior. To account for this, we provide power analyses for both scenarios–e.g., responses about oneself, and responses about doctors.

This yields twelve pairwise comparisons: six for responses based on expectations of surgeon behavior, and six for responses based on oneself. Figure 3 shows the power analyses for pairwise comparisons when means based on survey responses regarding surgeons. Figure 4 shows the power analyses when the means are based on survey responses regarding oneself.

When expected means are based on responses regarding surgeons, we would have near, or greater than, 80% power to detect differences in means with a sample of 100 doctors. The X-axis shows the sample for the given comparison. There are four experimental groups covering approximately 100 doctors, and so half of the sample (50) is used for a given pairwise comparison. We would achieve this sample size with the department of surgery at UUHC. As noted earlier, we are under agreement to perform the same experiment at University of

Chicago Medicine. UUHC'S Orthopedics department has also solicited a proposal from us to apply the experiment there. This would yield a sample size of approximately 300 surgeons across UUHC and University of Chicago Medicine.

A sample of that size would provide near, or greater than 80% power for all pairwise comparisons when expected means are based on responses regarding surgeons. It would provide near, or greater than 80% power for four of six pairwise comparisons when expected means are based on responses regarding oneself. We have expanded from UUHC to Duke Surgery. We are awaiting data from UChicago Medicine to expand there. Those collectively would increase our sample to roughly 400 doctors and provide sufficient power for the remaining two pairwise comparisons: information alone versus comparison, and comparison versus altruism.

5.4 Randomization

We implemented a block randomization algorithm, in which randomization was blocked on division (e.g. the Cardiothoracic Surgery division within the Surgery department), surgeon gender, and savings opportunity (i.e., above or below median). This generates balance in treatments by each of these features. Because our pre-survey provided evidence that all treatments had the potential to produce treatment effects, sample sizes were distributed evenly across the four treatment arms during the first wave of the experiment.

5.5 Rollout (Preliminary)

We are rolling out the study across surgeons in the department of surgery at UUHC and at Duke Ambulatory Surgery Center. As stated in the introduction, as a final quality check before rollout at UUHC and Duke Surgery, surgery department leadership reviewed each suggested DPC change and approved before sending. They also reviewed the experimental treatment messaging to ensure the language was appropriate, and approved the randomized messaging design.

The emails with DPC change suggestions and treatment messaging come from a surgeon in department leadership (e.g., surgery department chair), and the specific wording used in the emails is shown in Figure 5. Although we distribute these emails using Qualtrics, the emails come from the surgery department chair's email address. This helps with take-up by making the request credible to the participants. The use of Qualtrics allows surgeons to click a button in the email to record their response to the suggestion.

Each email shows the surgeon a suggested change that would result in a DPC that has been used at the field site for the same procedure. The surgeon has three options to respond to the suggestion, embedded as buttons to click in the email: 1) “Accept All,” 2) “Select Drops and Additions,” or 3) “Reject All.” If a surgeon clicks “Select Drops and Additions,” the surgeon will be taken to a Qualtrics survey with the following sequence. First, they will see the list of the suggested tools to drop, where they can “deselect” any tools that they wish to keep on their DPC. Second, they will see the list of all other tools on their DPC, and they can choose to drop any *additional* items, beyond what we have suggested. Finally, they will see the list of the suggested tools to add, and they can modify these suggested additions. In each list, we will show the per-unit cost of each supply so that surgeons are aware of how much their drops save and how much their additions cost.

Our main dependent variable of interest from the field experiment is *Degree Savings*, or the degree to which the surgeon agreed to savings of the amount offered. If the surgeon chose "Accept All," that variable takes a value of 1. If the surgeon chose "Reject All," that variable takes a value of 0. If the surgeon opted to select their own drops and additions, we calculate the savings from their choices as a portion of the total *suggested* savings. Thus, *Degree Savings* could take a value of greater than one, if the surgeon makes additional drops that total to more than the suggested amount, or less than one, if the surgeon accepts a subset of the suggested drops. In subsequent months, we will collect *actual* savings from our intervention (e.g., total supply cost spend), though because we have automated the link between the “accept” button and the DPC change, actual savings should naturally follow.

After we provide the first suggestion to a given surgeon, we will send another email within two weeks with a second suggestion, and so on, until we have made at least four suggestions to each surgeon. The treatment conditions will stay constant by surgeon during the experiment. The resulting panel data, paired with granular data on swap characteristics and surgeon characteristics, will allow tests of the conditions under which swaps are more likely to be accepted and under which certain treatments work best. Based on our Prolific survey findings, both our treatments as well as characteristics such as gender seem likely to play a significant role.

We have received results from applying the experiment to seventeen surgeons in the general surgery division at UUHC and fourteen surgeons in the general surgery division at Duke. The majority of surgeons accepted all of the suggestions we made, and the projected savings was $257,009 per year across seventeen surgeons for an average savings of $15,118 per surgeon per year. We will conduct statistical analyses comparing the treatment condition effects after distributing the suggestions to a larger sample.

6 Conclusion

Despite the magnitude of healthcare costs, policymakers and hospital administrators struggle to implement effective strategies to curb its rapid growth. In this paper, we use a field experiment in combination with a detailed dataset of hospital supplies and costs provided by UUHC to investigate whether interventions that provide doctors with information about cost-effective supply options causally lead to decreased healthcare costs. We supplement our information intervention with three types of incentives (altruism, peer comparison, and financial incentives) validated by a Prolific survey conducted before the implementation of our field experiment.

Effectively testing whether or not our interventions lead to lower healthcare spending without sacrificing quality requires three steps. First, we validated that lower costs are not correlated with quality in the supply options that we identify and present to doctors. We validated this through an elastic net model as well as an analysis of stock outs. As an extra safeguard, potential swaps were also audited by the hospital. Second, we validated our information and incentive interventions through a Prolific survey. The pre-experimental survey allows us to ensure that our treatments are well motivated and have the potential to influence doctor behavior. The final step is the actual implementation of the field experiment. We are currently rolling out the field experiment at UUHC and Duke surgery. In the initial rollout, the majority of surgeons accepted the suggestions yielding a projected $15,118 annual savings per surgeon. We expect comprehensive results by the end of Summer 2025.

References

ADIDA, E., H. MAMANI, AND S. NASSIRI (2017): "Bundled payment vs. fee-for-service: Impact of payment scheme on performance," *Management Science*, 63, 1606–1624.

AGHION, P. AND J. TIROLE (1997): "Formal and real authority in organizations," *Journal of political economy*, 105, 1–29.

ANDREONI, J. (1989): "Giving with impure altruism: Applications to charity and Ricardian equivalence," *Journal of political Economy*, 97, 1447–1458.

——— (1990): "Impure altruism and donations to public goods: A theory of warm-glow giving," *The economic journal*, 100, 464–477.

BECKER, G. S. (1974): "A theory of social interactions," *Journal of political economy*, 82, 1063–1093.

BERRY JAEKER, J. A. AND A. L. TUCKER (2017): "Past the point of speeding up: The negative effects of workload saturation on efficiency and patient severity," *Management Science*, 63, 1042–1062.

DEKONENKO, C., T. A. OYETUNJI, AND R. M. RENTEA (2020): "Surgical tray reduction for cost saving in pediatric surgical cases: A qualitative systematic review," *Journal of pediatric surgery*, 55, 2435–2441.

DELLAVIGNA, S., J. A. LIST, U. MALMENDIER, AND G. RAO (2022): "Estimating social preferences and gift exchange at work," *American Economic Review*, 112, 1038–1074.

ERAT, S. AND U. GNEEZY (2016): "Incentives for creativity," *Experimental Economics*, 19, 269–280.

FERRARO, P. J. AND M. K. PRICE (2013): "Using nonpecuniary strategies to influence behavior: Evidence from a large-scale field experiment," *Review of Economics and Statistics*, 95, 64–73.

GNEEZY, U., S. MEIER, AND P. REY-BIEL (2011): "When and why incentives (don't) work to modify behavior," *Journal of economic perspectives*, 25, 191–210.

GNEEZY, U., M. NIEDERLE, AND A. RUSTICHINI (2003): "Performance in competitive environments: Gender differences," *The quarterly journal of economics*, 118, 1049–1074.

GOSNELL, G. K., J. A. LIST, AND R. D. METCALFE (2020): "The impact of management practices on employee productivity: A field experiment with airline captains," *Journal of Political Economy*, 128, 1195–1233.

HALLSWORTH, M., J. A. LIST, R. D. METCALFE, AND I. VLAEV (2017): "The behavioralist as tax collector: Using natural field experiments to enhance tax compliance," *Journal of public economics*, 148, 14–31.

HARRIS, M. AND A. RAVIV (2002): "Organization design," *Management science*, 48, 852–865.

JAMALABADI, S., V. WINTER, AND J. SCHREYÖGG (2020): "A systematic review of the association between hospital cost/price and the quality of care," *Applied Health Economics and Health Policy*, 18, 625–639.

JANHOFER, D. E., C. LAKHIANI, AND D. H. SONG (2019): "Addressing surgeon fatigue: current understanding and strategies for mitigation," *Plastic and reconstructive surgery*, 144, 693e–699e.

KAHNEMAN, D. (1973): *Attention and effort*, vol. 1063, Citeseer.

KC, D. S. AND C. TERWIESCH (2009): "Impact of workload on service time and patient safety: An econometric analysis of hospital operations," *Management science*, 55, 1486–1498.

Maćkowiak, B., F. Matějka, and M. Wiederholt (2023): "Rational inattention: A review," *Journal of Economic Literature*, 61, 226–273.

Meckling, W. H. and M. C. Jensen (1976): "Theory of the Firm," *Managerial Behavior, Agency Costs and Ownership Structure*.

Porter, M. E. (2008): "Value-based health care delivery," *Annals of surgery*, 248, 503–509.

Scales Jr, C. D., T. Moin, A. Fink, S. H. Berry, N. Afsar-Manesh, C. M. Mangione, and B. P. Kerfoot (2016): "A randomized, controlled trial of team-based competition to increase learner participation in quality-improvement education," *International Journal for Quality in Health Care*, 28, 227–232.

Sims, C. A. (2003): "Implications of rational inattention," *Journal of monetary Economics*, 50, 665–690.

Skarda, D. E., M. Rollins, S. Andrews, M. McFadden, D. Barnhart, R. Meyers, and E. Scaife (2015): "One hospital, one appendectomy: the cost effectiveness of a standardized doctor's preference card," *Journal of pediatric surgery*, 50, 919–922.

Stein, J. C. (2002): "Information production and capital allocation: Decentralized versus hierarchical firms," *The journal of finance*, 57, 1891–1921.

Teisberg, E., S. Wallace, and S. O'Hara (2020): "Defining and implementing value-based health care: a strategic framework," *Academic Medicine*, 95, 682–685.

Zhang, P., Y. Zhang, and M. A. Palma (2024): "Social roles and competitiveness: My willingness to compete depends on who I am (supposed to be)," *Games and Economic Behavior*, 143, 125–151.

Figure 1. Example of a Potential Supply Swap
This figure presents examples of two catheters produced by different manufacturers. Both items are catheters with balloons for predilatation and stent placement in a percutaneous transluminal coronary angioplasty (PTCA). The specifications are 1.5mm x 20mm balloon size.

Terumo's Takeru PTCA Dialation Catheter
1.5 mm balloon * 20mm

COST: $75

Boston Scientific PTCA Dialation Catheter
1.5 mm balloon * 20mm

COST: $40

Figure 2. Geographic Dispersion of Survey Respondents
This figure illustrates the geographic distribution of respondents of the Prolific survey. Locations are noted in Prolific with latitudes and longitudes.

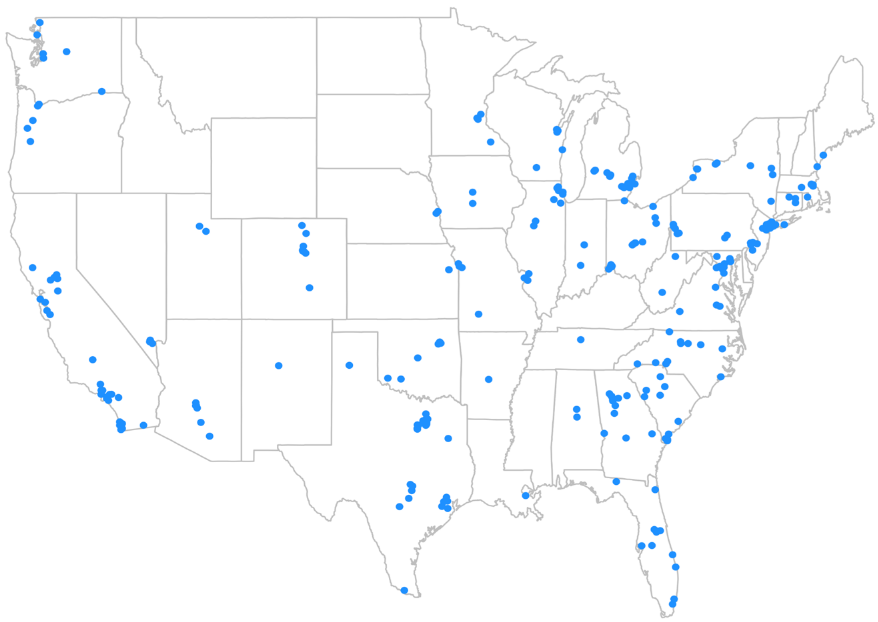

Figure 3. Power Analyses with Means from Survey Regarding Surgeons

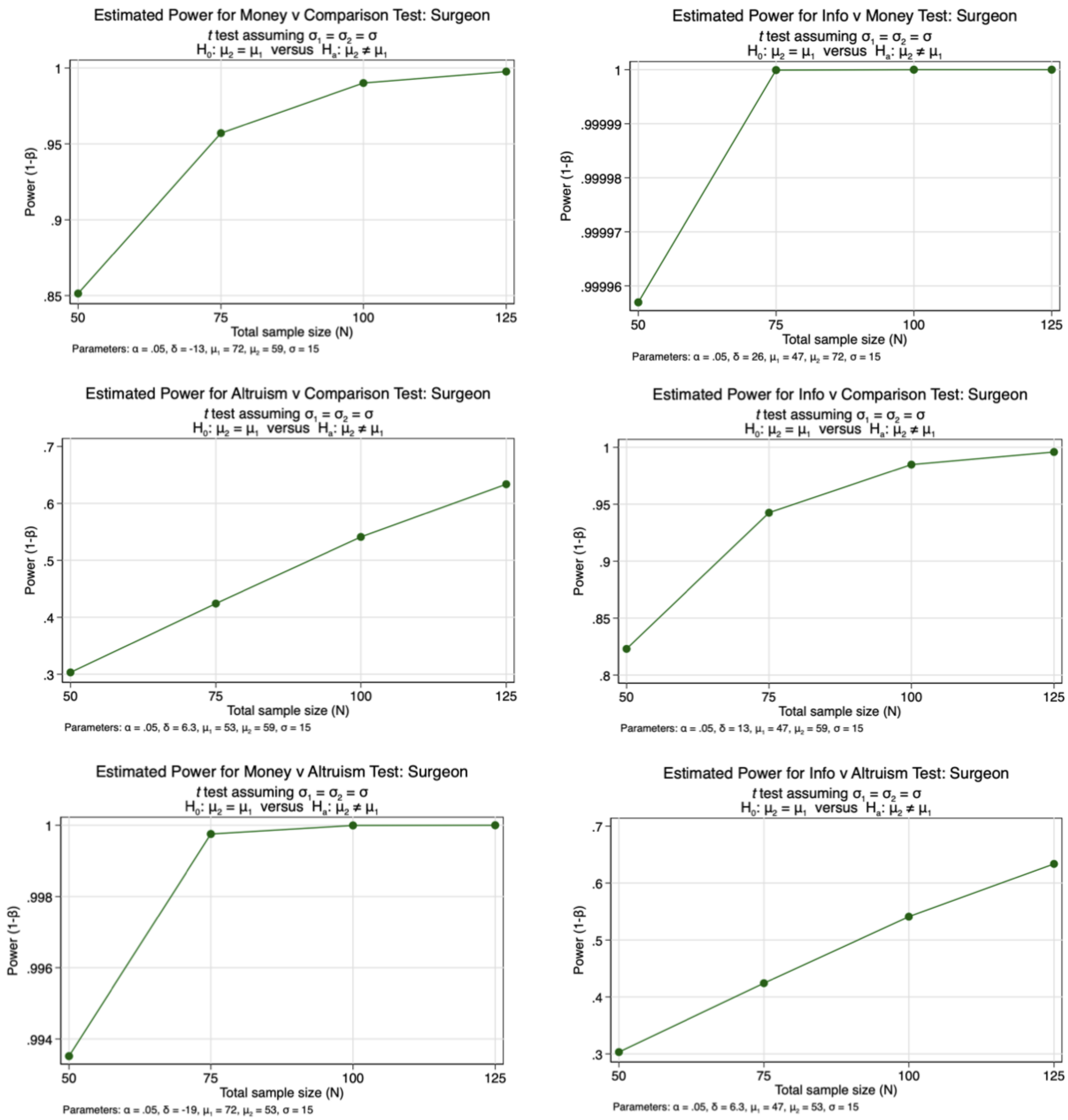

Figure 4. Power Analyses with Means from Survey Regarding Self

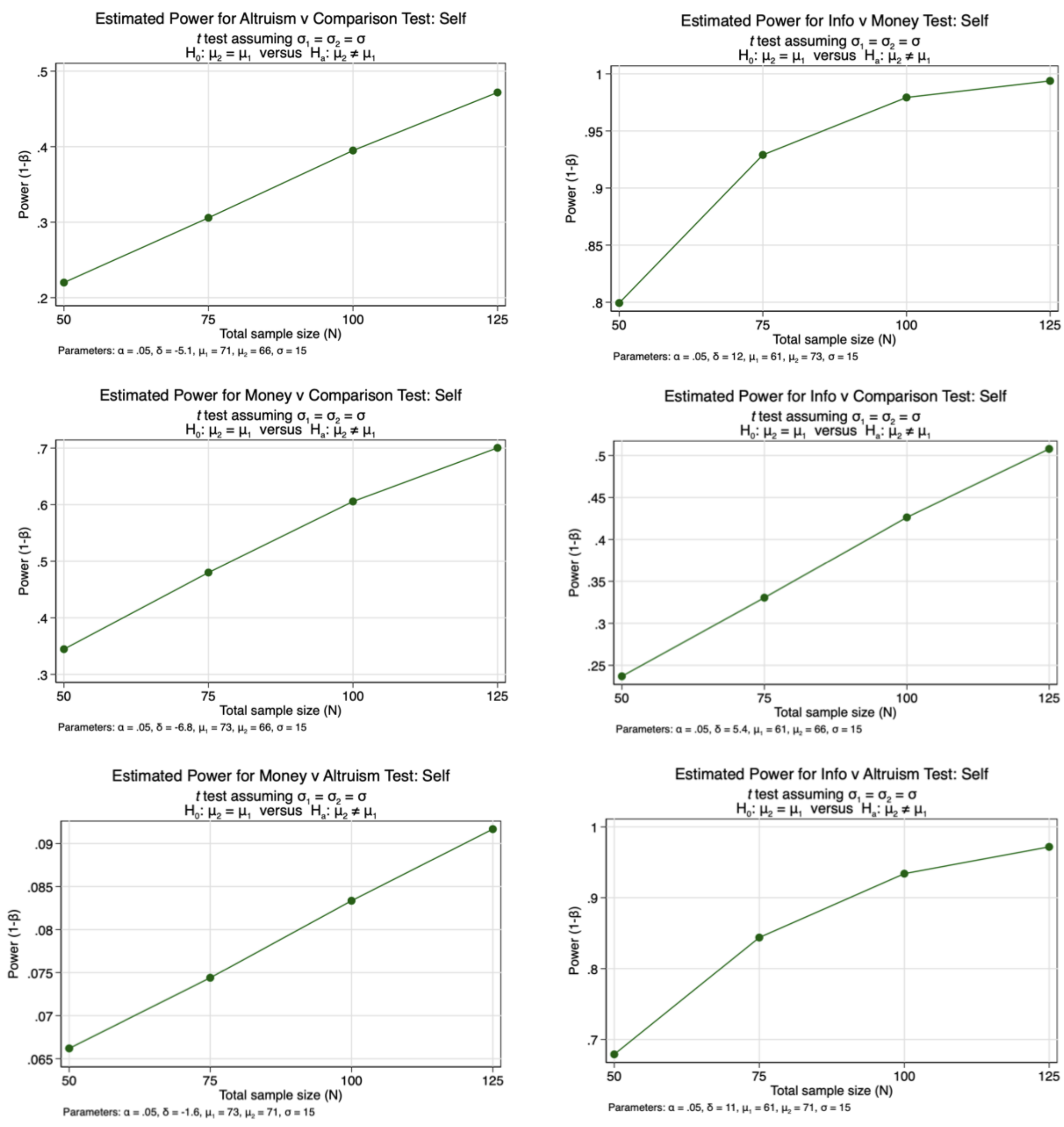

Figure 5. Treatment Language

This figure presents the messaging used in our three treatment groups, appealing to altruism, peer comparison, and financial incentives.

Altruism

As a reminder, patients pay a portion of the total cost of <procedure_name>. **By agreeing to make the changes below, you would pass an estimated savings of $x on to your patients each year by lowering their out-of-pocket costs**.

Peer Comparison

We performed a cost analysis of all surgeons performing <procedure_name>, and found that **____ (e.g, 9) out of ____ (e.g., 12) doctors who do this procedure at UUHC have a lower cost DPC than yours.**

Financial Incentives

We recognize that your effort in cost reduction is beneficial to the hospital and would like to share that benefit with you. **By agreeing to make the changes below, we will pay 30% of the first year's savings that you generate into your division's budget.**

Table 1: Descriptive Statistics from Survey
This table reports the descriptive statistics for the key demographic traits of Prolific survey respondents.

	N	Mean	Std. Dev	Min	Max
Age	286	36.028	11.354	20	75
Income	286	73,237.439	47,837.852	0	250,000
Female	286	0.644			
Doctorate	286	0.091			
Masters	286	0.315			
Bachelors	286	0.787			
High School	286	0.888			

Table 2: Survey Responses

This table reports the responses to the Prolific survey. Responses are recorded for two blocks: one block asks whether the respondent would personally agree to the supply switch under each treatment condition (noted in the table as *Self_Mean*), and it asks the respondent to weight their level of confidence in their answer (noted in the table as *Self_Conf.*); the second block asks the respondent to estimate how likely a surgeon would be to agree to the switch under each treatment condition (noted in the table as *Surgeon_Mean*), and it asks the respondent to weight their level of confidence in their answer (noted in the table as *Surgeon_Conf.*). The blocks are randomized across respondents. Mean responses are recorded on a 7 point scale, and for each question, we ask the respondent to rate their confidence in their response on a 100 point scale. Standard deviations are reported in parentheses. In Panel A, we report the mean values and standard deviations of each response. In Panel B, we report the differences across group means. P-values are noted and *** represents significance at 0.01. Significance is assessed using the Wilcoxon signed-rank test.

(a) Panel A: Mean and Standard Deviations of Survey Responses

Treatment	Self_Mean	Self_Conf.	Surgeon_Mean	Surgeon_Conf.
Control (Info. Only)	4.839	78.5	3.783	68.213
	(1.495)	(20.754)	(1.515)	(21.461)
Altruism	5.678	82.283	4.231	68.846
	(1.320)	(20.801)	(1.497)	(20.224)
Peer Comparison	5.259	79.717	4.734	69.605
	(1.539)	(21.294)	(1.386)	(19.755)
Financial Incentives	5.615	82.594	5.804	76.112
	(1.479)	(19.987)	(1.221)	(19.677)

(b) Panel B: Differences in group means

	Altr._Self	Comp._Self	Fin._Self	Control_Self	Altr._Surg.	Comp._Surg.	Fin._Surg.	Control_Surg.
Altruism_Self	0							
Comparison_Self	-0.42***	0						
Financial_Self	-0.06	0.36***	0					
Control_Self	-0.84***	-0.42***	-0.78***	0				
Altruism_Surgeon	-1.45***	-1.03***	-1.38***	-0.61***	0			
Comparison_Surgeon	-0.94***	-0.52***	-0.88***	-0.1	0.5***	0		
Financial_Surgeon	0.13	0.55***	0.19	0.97***	1.57***	1.07***	0	
Control_Surgeon	-1.9***	-1.48***	-1.83***	-1.06***	-0.45***	-0.95***	-2.02***	0

Table 3: Survey Responses: Split by the Gender of the Respondent
This table reports the responses to the Prolific survey, split by the gender of the survey respondent. Responses are recorded for two blocks: one block asks whether the respondent would personally agree to the supply switch under each treatment condition (noted in the table as *Self_ Mean*), and it asks the respondent to weight their level of confidence in their answer (noted in the table as *Self_ Conf.*); the second block asks the respondent to estimate how likely a surgeon would be to agree to the switch under each treatment condition (noted in the table as *Surgeon_ Mean*), and it asks the respondent to weight their level of confidence in their answer (noted in the table as *Surgeon_ Conf.*). The blocks are randomized across respondents. Mean responses are recorded on a 7 point scale, and for each question, we ask the respondent to rate their confidence in their response on a 100 point scale. Standard deviations are reported in parentheses.

	Female Respondent			
Treatment	Self_Mean	Self_Conf.Mean	Surgeon_Mean	Surgeon_Conf.
Control (Info. Only)	5.071	79.503	3.820	68.410
	(1.483)	(21.196)	(1.499)	(22.318)
Altruism	5.836	83.454	4.268	68.984
	(1.295)	(21.131)	(1.572)	(21.097)
Peer Comparison	5.497	81.612	4.710	69.164
	(1.486)	(21.204)	(1.402)	(19.648)
Financial Incentives	5.563	81.279	5.885	75.169
	(1.481)	(21.257)	(1.183)	(20.527)

	Male Respondent			
Treatment	Self_Mean	Self_Conf.Mean	Surgeon_Mean	Surgeon_Conf.
Control (Info. Only)	4.406	76.495	3.713	67.525
	(1.437)	(20.031)	(1.558)	(19.947)
Altruism	5.376	79.861	4.149	68.960
	(1.326)	(20.182)	(1.359)	(18.627)
Peer Comparison	4.822	76.277	4.782	70.634
	(1.545)	(21.155)	(1.361)	(19.994)
Financial Incentives	5.693	84.663	5.634	77.960
	(1.488)	(17.423)	(1.278)	(18.073)

Table 4: Descriptive Statistics: Patient-level Controls and Outcomes at UUHC Surgery

This table reports the descriptive statistics of the procedure data and outcomes at UUHC Surgery. *BMI* is the patient's body mass index; *CCI score* is the Charlson Comorbidity Index, which predicts the ten-year mortality for a patient who may have a range of comorbid conditions; *Age* is the age of the patient in years; *Mortality* is set equal to one if the procedure resulted in patient death; *30-day readmission* is set equal to one if the patient was readmitted to the hospital within 30 days of the procedure; *Surgery time* is the time, in minutes, from the start of the procedure to the end; *Length of stay* is the number of days, post-surgery, that the patient stayed in the hospital; and *Tool types per DPC* is the number of unique tool types included in the doctor's DPC card for each procedure.

	N	Mean	Std. Dev.	Min.	25%	Median	75%	Max.
BMI	104,912	29.13	7.54	13.54	23.99	27.90	32.83	79.47
CCI score	125,195	2.59	2.89	0	0	2	4	15
Age	129,189	50.93	18.40	0	37	53	66	90
Mortality	45,174	0.04	0.20	0	0	0	0	1
30-day readmission	49,220	0.05	0.22	0	0	0	0	1
Surgery time	129,189	107.82	102.51	2	40	77	139	757
Length of stay post-surgery	90,517	3.72	9.99	0.00	0.08	0.25	3.10	140.59
Types of tools in DPC	129,189	27.51	15.11	1	16	25	37	151

Table 5: Descriptive Statistics: Patient-level Controls and Outcomes at Duke Ambulatory Surgery Center

This table reports the descriptive statistics of the procedure data and outcomes at Duke Ambulatory Surgery Center. *BMI* is the patient's body mass index; *Mortality risk* is an index for mortality risk associated with the patient's underlying health, ranging from 0 for no additional risk due to underlying health, to 4 indicating extreme risk; *Age* is the age of the patient in years; *Mortality* is set equal to one if the procedure resulted in patient death; *30-day readmission* is set equal to one if the patient was readmitted to the hospital within 30 days of the procedure; *Length of stay* is the number of days, post-surgery, that the patient stayed in the hospital; and *Tool types per DPC* is the number of unique tool types included in the doctor's DPC card for each procedure.

	N	Mean	Std. Dev.	Min.	25%	Median	75%	Max.
BMI	7,302	30.28	7.65	17.06	25.11	27.90	33.97	54.81
Mortality risk	7,553	0.48	0.93	0	0	0	1	4
Age	129,189	50.93	18.40	0	37	53	66	90
Mortality	2,096	0.02	0.12	0	0	0	0	1
30-day readmission	7,059	0.04	0.22	0	0	0	0	1
Length of stay post-surgery	7,553	1.50	4.71	0.00	0.00	0.00	1.00	16.00
Tool types per DPC	7,570	39.77	15.28	1	30	37	47	144

Table 6: The Effect of Individual Supplies on Patient Outcomes
This table reports the results from a logistic regression with an elastic net regularization penalty. The L1 ratio was set at 0.9, and the regularization strength parameter, C, was set to 0.001. The model's hyperparameters were optimized using GridSearchCV with Related Stratified K-Folds cross-validation. Model selection was assessed using the area under the receiver operating characteristic (ROC). The resulting analysis, below, reports only the variables that had a significant impact on mortality. *MS-DRG* is the Medicare Severity-Diagnosis Related Group; higher weights indicate more patient risk. *Patient Age* is the patient's age in years; *CCI Score* is the Charlson Comorbidity Index, a method of categorizing comorbidities of patients based on the International Classification of Diseases (ICD) diagnosis codes, which are associated with the risk of death within one year of hospitalization. Coefficients, odds ratios, and standard errors are reported for each variable. In Panel A, the dependent variable is patient mortality (set equal to one in the event of patient death). In Panel B, the dependent variable is the patient's 30-day readmission (set equal to one in the event of readmission to the hospital within a 30 day period).

(a) Panel A: Mortality

	Mortality		
	Coefficient	Odds Ratio	Std. Error
MS-DRG weight	0.7393	2.0945	0.0006
Patient Age	0.2138	1.2383	0.0005
CCI Score	0.1706	1.1860	0.0005
Drape 89124	0.13	1.14	0.00
Specimen Supplies 83050	0.12	1.12	0.00
Suction 75495*	-0.06	0.94	0.00
Drape 72128	0.06	1.06	0.00
Staple 72338*	0.05	1.06	0.00
Implant 85953	-0.05	0.95	0.00
Dressing 83330	-0.05	0.95	0.00
Drape 89020*	-0.03	0.98	0.00
Dressing 88233	-0.02	0.98	0.00
Airway Trach 80735	0.02	1.02	0.00
Dressing 83018	0.02	1.02	0.00
Endoscopy 86161	0.02	1.02	0.00
Tube 97853*	0.02	1.02	0.00
Endoscopy 86575	-0.01	0.99	0.00
Tube 83058*	-0.01	0.99	0.00
Catheter Unigrip 183121	0.01	1.01	0.00
Dressing 83341	0.00	1.00	0.00
Implant 86884	-0.00	1.00	0.00

(b) Panel B: 30-day Readmission

	30-day Readmission		
	Coefficient	Odds Ratio	Std. Error
CCI score	0.1675	1.1823	0.0005
BMI risk	-0.0333	0.9673	0.0005
Patient Age	-0.0118	0.9882	0.0005
Pack 83356	0.12	1.12	0.0004
Drape 89020*	-0.08	0.92	0.0004
Dressing 90356*	-0.06	0.94	0.0004
Laser 97880	0.06	1.06	0.0003
Endoscopy 86575	-0.05	0.95	0.0004
Clip 84895	-0.05	0.95	0.0006
Dressing 88233	-0.04	0.96	0.0004
Drape 81544*	-0.04	0.96	0.0004
Drape 89017	0.04	1.04	0.0004
Tube 84749	-0.04	0.96	0.0004
Tube 83058*	-0.04	0.96	0.0004
Dressing 105977	0.03	1.03	0.0004
Syringe/Needle 90214	0.02	1.02	0.0003
Miscellaneous 82309	0.02	1.02	0.0004
Drape 81772	-0.02	0.98	0.0005
Stapler 72260*	0.02	1.02	0.0020
Single Action Pump 98025	0.01	1.01	0.0003
Implant 90353	0.01	1.01	0.0003
Solution 84740	-0.01	0.99	0.0006
Endoscopy 81656	0.01	1.01	0.0005
Staple 72271*	0.01	1.01	0.0020
Endoscopy 88630	0.01	1.01	0.0005
Syringe/Needle 84834	0.01	1.01	0.0003
Cautery 77425*	-0.00	0.99	0.0004
Endoscopy 87963	-0.00	1.00	0.0004
Implant 83137	0.00	1.00	0.0003
Tube 86258	-0.00	1.00	0.0004
Solution 85548	-0.00	1.00	0.0005
Clamp 82179	0.00	1.00	0.0006

Table 7: The Effect of Individual Supplies on Patient Outcomes: Evidence from Stock Outs
This table reports the results from estimating Equation 1. We identify stock outs as follows: supplies must be in use for a given period by two or more doctors performing the same primary procedure. We then require that the supply disappear from the DPC of all doctors, for a minimum of 30 days (during which time the doctors continued to perform that same procedure). Finally, we require that the same item appears back on the DPCs at some later date (after the stock-out period). For those procedures with stock-outs, we create a continuous variable, *Stockout_Dollar_Change* which represents the change in the price of the doctor's DPC from the period during the stock out, relative to the period before the supply was stocked out. For procedures without stock outs, this variable is zero. The regressions includes doctor, quarter, and procedure fixed effects, and standard errors are clustered at the procedure level. Column 1 presents the results where the dependent variable is mortality, and column 2 presents the results where the dependent variable is 30-day readmissions. ***, **, * represent significance at the 0.01, 0.05, and 0.1 levels, respectively.

	Mortality	30-day Readmission
Stockout_Dollar_Change	0.000	-0.000
	(0.000)	(0.000)
CCI	0.004***	0.004***
	(0.001)	(0.001)
Age	0.001***	-0.000***
	(0.000)	(0.000)
MS-DRGt	0.011***	-0.000
	(0.001)	(0.000)
N	43,702	47,231
Adj. R^2	0.190	0.046
ProcedureFE	Y	Y
QuarterFE	Y	Y
DoctorFE	Y	Y

Table 8: DPC Kit Suggestions at UUHC - Descriptives

This table provides the descriptive statistics of the set of DPC swaps that will be suggested in our experiment. *Savings per use* is calculated as the total amount of cost savings realized if the doctor switched from her current DPC to our suggested DPC. *Yearly savings* is calculated as the savings per use multiplied by the total number of times the procedure was performed in 2023. *Total tool changes* is the total number of supplies recommended for the change – including both dropped supplies and added supplies.

	Mean	Std. Dev.	Min	25%	Median	75%	Max
Yearly savings	49,494.23	62,295.26	12,029.40	16,654.21	34,004.43	40,203.49	272,430.40
Savings per use	1,854.12	2,836.73	295.94	949.23	1,138.80	1,397.69	13,621.52
Total Tool Changes	5.55	1.99	2.00	4.75	5.50	7.00	8.00
Unique Tools added	0.85	1.04	0.00	0.00	1.00	1.00	4.00
Unique Tools removed	4.50	1.88	2.00	3.75	4.00	6.00	8.00

Table 9: DPC Kit Suggestions at Duke Ambulatory Surgery Center - Descriptives
This table provides the descriptive statistics of the set of DPC swaps that will be suggested in our experiment. *Savings per use* is calculated as the total amount of cost savings realized if the doctor switched from her current DPC to our suggested DPC. *Yearly savings* is calculated as the savings per use multiplied by the total number of times the procedure was performed in 2023. *Total tool changes* is the total number of supplies recommended for the change – including both dropped supplies and added supplies.

	Mean	Std. Dev.	Min	25%	Median	75%	Max
Yearly savings	26,834.30	35,030.03	570.00	4,893.79	10,137.88	34,220.38	157,177.60
Savings per use	667.00	833.85	23.15	195.41	436.86	828.80	5,991.62
Total Tool Changes	5.05	2.16	1.00	4.00	5.00	7.00	8.00
Unique Tools added	0.56	0.89	0.00	0.00	0.00	1.00	4.00
Unique Tools removed	4.49	2.12	1.00	3.00	4.00	6.00	8.00

7 Appendix

A1. Prolific Survey

This appendix contains the main content of the survey we conducted on Prolific, using respondents from the field of healthcare.

Please read the following description and proceed to answer the questions that follow.

Hospital A is implementing a quality improvement initiative to decrease operating room supply costs. Before surgeons perform a procedure, they choose which supplies to use. However, surgeons often select expensive supplies, despite the fact that there are many less expensive options that produce the same quality outcomes. This has resulted in increased costs for the hospital, and many of these costs are passed on to the patient.

The hospital is hoping to improve cost-efficiency in their supply management by implementing an intervention to encourage doctors to reconsider their supply choices. Importantly, the intervention is aimed at reducing costs for common goods (e.g., bandages, towels, catheters, etc.) that will not fundamentally change the way the doctor performs the procedure. Moreover, all supplies that are recommended have been verified to have “high quality” in terms of patient outcomes. The intervention will involve (1) providing doctors with the option to easily switch supplies to a cheaper supply kit that is of equal or higher quality to their existing supplies and (2) providing incentives to each doctor to encourage them to agree to switch their supplies.

Which of the following is part of the quality improvement intervention at Hospital A? (choose all options that apply)

- [ ] Providing doctors with the option to easily switch supplies to a cheaper supply kit that is of equal or higher quality to their existing supplies.
- [ ] Providing incentives to each doctor to encourage them to agree to switch their supplies.
- [ ] Providing hospital administrators with training to better utilize their cost management software.

For each of the prompts below, please answer how likely you think **A SURGEON** who is performing the procedure would be to switch their supplies.

Notes: We are not asking how you would respond, but rather we are asking you to predict how a surgeon would respond. Please keep in mind that surgeons are very time-constrained and thus inducing them to make supply changes might be difficult.

(a) On a scale of 1-7, how likely would a **surgeon** be willing to switch to lower cost supplies if we told them that **a portion of their cost savings would be passed on to the patient?**

Extremely unlikely	Somewhat unlikely		Neither likely nor unlikely	Somewhat likely		Extremely likely
1	2	3	4	5	6	7

(b) On a scale of 1-100 (with 100 being absolute certainty), how confident are you in your answer to (a)? Please assess the likelihood that the true answer falls between +1 and -1 of your response.

E.g., if you responded '4', what is the chance that the true likelihood with between 3 and 5?

0 10 20 30 40 50 60 70 80 90 100

(c) On a scale of 1-7, how likely would a surgeon be willing to switch to lower cost supplies if we provided a peer comparison to their supply costs, and **we showed them that their spending exceeded that of peer doctors performing the same procedure?**

Extremely unlikely	Somewhat unlikely		Neither likely nor unlikely	Somewhat likely		Extremely likely
1	2	3	4	5	6	7

(d) On a scale of 1-100 (with 100 being absolute certainty), how confident are you in your answer to (c)? Please assess the likelihood that the true answer falls between +1 and -1 of your response.

0 10 20 30 40 50 60 70 80 90 100

(e) On a scale of 1-7, how likely would a surgeon be willing to switch to lower cost supplies if **a portion of their cost savings would be kept for themselves** (deposited in the surgeon's departmental research budget)?

Extremely unlikely	Somewhat unlikely		Neither likely nor unlikely	Somewhat likely		Extremely likely
1	2	3	4	5	6	7

(f) On a scale of 1-100 (with 100 being absolute certainty), how confident are you in your answer to (e)? Please assess the likelihood that the true answer falls between +1 and -1 of your response.

0 10 20 30 40 50 60 70 80 90 100

(g) On a scale of 1-7, how likely would a surgeon be willing to switch to lower cost supplies if we simply provided them with **information** about alternative supplies that were both cheaper and of equivalent quality to the ones they are currently using, but do not provide them with additional **incentives** to switch?

Extremely unlikely	Somewhat unlikely		Neither likely nor unlikely	Somewhat likely		Extremely likely
1	2	3	4	5	6	7

(h) On a scale of 1-100 (with 100 being absolute certainty), how confident are you in your answer to (g)? Please assess the likelihood that the true answer falls between +1 and -1 of your response.

0 10 20 30 40 50 60 70 80 90 100

Please explain the reasoning behind your highest ranked choice on this page.

Powered by Qualtrics

For each of the prompts below, please imagine **YOU** are making the decision to switch the supplies as if **YOU** were performing the surgical procedure.

(a) On a scale of 1-7, how likely would you be willing to switch to lower cost supplies if we told you that **a portion of the cost savings would be passed on to the patient?**

Extremely unlikely		Somewhat unlikely	Neither likely nor unlikely	Somewhat likely		Extremely likely
1	2	3	4	5	6	7

(b) On a scale of 1-100 (with 100 being absolute certainty), how confident are you in your answer to (a)? Please assess the likelihood that the true answer falls between +1 and -1 of your response.

E.g., if you responded '4', what is the chance that the true likelihood with between 3 and 5?

0 10 20 30 40 50 60 70 80 90 100

(c) On a scale of 1-7, how likely would you be willing to switch to lower cost supplies if we provided a peer comparison to your supply costs, and **we showed you that your spending exceeded that of peer doctors performing the same procedure?**

Extremely unlikely		Somewhat unlikely	Neither likely nor unlikely	Somewhat likely		Extremely likely
1	2	3	4	5	6	7

(d) On a scale of 1-100 (with 100 being absolute certainty), how confident are you in your answer to (c)? Please assess the likelihood that the true answer falls between +1 and -1 of your response.

0 10 20 30 40 50 60 70 80 90 100

(e) On a scale of 1-7, how likely would you be willing to switch to lower cost supplies if **a portion of your cost savings would be kept for yourself** (deposited in your departmental research budget)?

Extremely unlikely		Somewhat unlikely	Neither likely nor unlikely	Somewhat likely		Extremely likely
1	2	3	4	5	6	7

(f) On a scale of 1-100 (with 100 being absolute certainty), how confident are you in your answer to (e)? Please assess the likelihood that the true answer falls between +1 and -1 of your response.

0 10 20 30 40 50 60 70 80 90 100

(g) On a scale of 1-7, how likely would you be willing to switch to lower cost supplies if we simply provided you with **information** about alternative supplies that were both cheaper and of equivalent quality to the ones you are currently using, but we did not provide you with additional **incentives** to switch?

Extremely unlikely	Somewhat unlikely		Neither likely nor unlikely	Somewhat likely		Extremely likely
1	2	3	4	5	6	7

(h) On a scale of 1-100 (with 100 being absolute certainty), how confident are you in your answer to (g)? Please assess the likelihood that the true answer falls between +1 and -1 of your response.

0 10 20 30 40 50 60 70 80 90 100

Please explain the reasoning behind your highest ranked choice on this page.

Powered by Qualtrics

A2. Common Supplies

This appendix contains the list of product categories that meet the criteria of "common, undifferentiated supplies," and which qualify for our set of supply swaps. The list was selected and annotated by doctors at UUHC and Duke Health.

Non-Differentiated, Common Supply Cateogries
Drape
Dressing
Solution
Syringe/needle
Endoscopy
Suction
Prep
Sponge
Supply
Tube
Staple
Cautery
Power
Catheter
Electrode
Vascular Access
Blade
Specimen Supplies
IV
Sheath
Laser
Pin
Personal Protection Equipment
Pack
Tubing
Cannula
Clip
Wire
Imaging
Airway
Needle
Cable
Sterile Processing
Dilator
Bur
Ablation Catheter
Biopsy Forcep
Syringe
Drain
Lead
Sizer
Lens