Focusing Provider Attention:
An Empirical Examination of Incentives and Feedback in Flu Vaccinations

Robert J. Niewoehner III*
University of North Carolina at Chapel Hill
Campus Box 3490, McColl Building
Chapel Hill, NC 27599-3490
robert_niewoehner@kenan-flagler.unc.edu

Bradley R. Staats
University of North Carolina at Chapel Hill
Campus Box 3490, McColl Building
Chapel Hill, NC 27599-3490
Tel: 919-962-7343
bstaats@unc.edu

15 August 2019

*** Corresponding Author**

Acknowledgments

We thank management at our research site for their cooperation. Any mistakes are our own.

Focusing Provider Attention:
An Empirical Examination of Incentives and Feedback in Flu Vaccinations

Abstract

The United States Center for Disease Control and Prevention classified the 2017-18 influenza (flu) season as the worst in recent history. Public health literature has estimated the average annual society costs for the flu in the US are $10.4 billion for direct medical costs and $16.3 billion in lost earnings. Despite these costs, the adult flu vaccination rate has remained between 37% and 44% each season between 2010 and 2018. Most research has concentrated on increasing the demand of patients seeking a vaccine with limited effect. An operations management approach would highlight the need to also consider the supply-side of the equation: driving compliance by focusing attention on the provision of the vaccine. Through the implementation of two interventions among individual health care clinics, we find this supply-side approach leads to an increase in the number of flu shots given. In 2017, the initial intervention in Florida was related to a 25% increase in flu shots. In 2018, a randomized, multi-state intervention led to a 12% increase in flu shots for the group provided performance feedback. Additionally, we compare the effects of introducing financial incentives and providing performance feedback. In the randomized, multi-state experiment, we find clinics who were provided performance feedback outperform the incentivized clinics in addition to outperforming the control group. Moreover, we also explore how performance feedback changes behavior, finding that clinics receiving performance feedback exhibit rank response behaviors, most notably Last Place Aversion. We discuss the implications of our work for healthcare operations, healthcare providers and healthcare administrators.

Key Words: Healthcare, Attention, Incentives, Feedback, Empirical Operations

1 Introduction

The United States Center for Disease Control and Prevention (CDC) classified the 2017-18 influenza (or flu) season as the worst in recent history (Garten et al. 2018, Figure 1). For 16 consecutive weeks, overall deaths attributed to influenza and pneumonia were at or above the epidemic threshold. Not only does influenza present significant health risks to large portions of the population, but it also leads to large economic costs through increased health care spending and missed days of work. Estimates for average annual society costs in the US are $10.4 billion for direct medical costs and $16.3 billion in lost earnings (Molinari et al. 2007). Although there is currently no method to completely eradicate the flu virus, experts continue to find that: “Receiving a seasonal influenza vaccine each year remains the best way to protect against seasonal influenza and its potentially severe consequences.” (Xu et al. 2019) In fact, the CDC recommends *everyone* over the age of 6 months be vaccinated every season.

Despite these recommendations, between 2010 and 2018, the adult flu vaccination rate remained between 37% and 44% each season (CDC 2019). Given this relatively low vaccination rate (compared to full compliance of 100%), it is not surprising that researchers in public health have focused a great deal on increasing vaccinations (Brewer et al. 2017). Most research in this area has concentrated on increasing the likelihood of patients seeking a vaccine – a “demand side” approach. The most successful of these interventions have engaged patients directly for example, by getting them to commit to a time (Milkman et al. 2011) or scheduling an appointment on their behalf (Chapman et al. 2016). A recent survey of the literature, however, finds that most interventions that have sought to alter patients’ thoughts and feelings regarding vaccinations have met with little apparent success (Brewer et al. 2017). Moreover, the same review notes the need for more field studies to design interventions that will work.

Thus, a quandary exists – the call to action for compliance is important due to the high potential health and economic costs from flu, however, existing approaches to address the gap have had limited success. Although a demand-focus is potentially a viable strategy, an operations management approach would *additionally* highlight the need to consider the supply-side of the equation: driving compliance by focusing attention on its provision. The question of compliance to standards is one that has been explored by the operations field from its earliest years (Taylor 1911) through today (Senot et al. 2016, Andritsos and Tang 2014, Corbett et al. 2005, Staats et al. 2017). Not only does the supply-side approach include ensuring that sufficient vaccine is produced and flows smoothly through the supply chain (Deo and Corbett 2009, Cho 2010, Arifoğlu et al. 2012), but also it highlights the key role that providers play in delivering a service. Healthcare operations frequently focuses on how providers shape a service system (e.g., Tucker 2007, Freeman et al. 2017) and recent work has explored interventions to alter provider behavior in positive ways (Tsai et al. 2015, Dai et al. 2017, Song and Tucker 2016, Song et al. 2018). In this paper we blend work on operational compliance and healthcare operations to investigate how to improve flu vaccination rates across the population by altering provider behavior.

Recent work examining the operational implications of individuals in service systems has highlighted the important role of individual discretion (e.g., van Donselaar et al. 2010, Campbell and Frei 2011, Kim et al. 2015, Phillips et al. 2015). Although sometimes discretion can lead to problematic outcomes (Ibanez et al. 2017, Ibanez and Toffel 2019) often it can improve outcomes (Freeman et al. 2017, Song et al. 2018). A key element comes in how the system is designed so that attention is focused on the critical elements of performance so that discretion is used productively. For example, Song et al. (2018) show that a small attention cue, like public performance feedback, can encourage low-performing physicians to seek out best practices and then improve their own processes. In this paper we consider how two attention cues: 1) incentives; and 2) feedback; may focus providers' attention on flu vaccination and lead to an increase in their delivery of the flu vaccine to patients, as compared to the prior year.

In total, our data includes over three million shot observations collected over four years from more than 1,400 clinics across 16 US states. We study the implementation of two flu vaccine incentive and performance feedback programs for individual health care clinics across multiple states in the US. We refer to these two interventions as Study 1 and Study 2. Study 1 occurred during Fall 2017 where 101 Florida clinics were recruited to participate; these clinics were provided both financial incentives for flu-shot growth and performance rankings relative to other intervention clinics. Building on the findings in this study, a field experiment was designed and Study 2 launched during Fall 2018, where 145 clinics from 9 different states were randomly sorted into three discrete groups: a control group which received no further intervention, a group which only received performance rankings based on flu shot growth, and a group which only received financial incentives based on flu shot growth. We find the introduction of attention cues led to an increase in flu shots for the treated clinics in both programs. In the latter study, we find that the clinics who receive performance feedback outperform the control group as well as the group which receives financial compensation based on growth. Surprisingly, the incentive group demonstrates the weakest Year-over-Year flu shot growth in Study 2.

Beyond exploring the increase in shots administered, our detailed data permits us to explore how the clinics respond to the program – in other words, not only whether the program leads to an overall increase in the desired behavior, but whether sub-groups benefit more (or less) by altering their behavior over time. The results from the first program indicated the possible presence of rank-response behaviors – in other words, that clinics appeared to change performance more in order to retain a high ranking or avoid a low ranking. Recent experimental work in the lab finds that individuals may exert extra effort to retain a high rank or avoid a particularly low rank (Kuziemko et al. 2014, Gill et al. 2019). Song et al.

(2018) find that when clinicians in a hospital receive public, relative performance feedback, the greatest performance improvement manifests among the poorest overall performers. We find that clinics near the ranking floor increase their vaccination effort seemingly to move away from last place. Even though these rankings are not incented, by themselves, they still meaningfully alter provider behavior, particularly amongst the clinics with the poorest performance. We demonstrate that, on average, clinics show improvement from the introduction of performance feedback, those ranked near the bottom show the greatest difference when compared with similar control clinics.

Our paper makes three primary contributions to the academic literature. First, we demonstrate that a novel, supply-side approach to vaccine management is operationally effective and yields an increase in provider flu vaccinations. This not only is an important practical result for healthcare, but it also adds to the burgeoning field of people-centric operations and the work on encouraging workers to use their discretion to improve operational performance. Additionally, these inventions occurred at the *clinic* rather than at the *provider* level; previous literature (e.g. Lazear 2000, Song et al. 2018) has primarily explored the impact of incentives and feedback at the *individual* level. As such, our work also demonstrates the effectiveness of intervening at the *firm* level, where treatment then diffuses down to lower, individual levels. Second, we demonstrate a real-world empirical context where performance feedback dominates the effect of financial incentives. Finally, in the context of a real-world setting, we test for and demonstrate the presence of rank-response behaviors, specifically Last-Place Aversion. We believe that we are the first to show that this is a dynamic effect, whereby effort is exerted to move out of last place each week. Overall, we find that private, relative performance feedback, can be an effective driver for operational change.

2 Hypothesis Development

The matching of supply to demand is a key tenet of operations management. We consider both aspects of supply and demand in our setting of flu vaccinations. Public health literature offers suggestions on how to increase demand for flu vaccinations and Brewer et al. (2017) identify three approaches for doing so: influencing patient thoughts and feelings surrounding vaccinations, leveraging a patient's social context, or changing behavior directly (e.g., using incentives or sanctions). Due largely to the lack of interventions robustly demonstrating an actual behavioral change, Brewer et al. (2017) identify the final approach as the most likely to affect actual vaccination behavior rather than simply altering beliefs or intentions. Successful studies include introducing prompts compelling employees to specify a date and time to receive their annual vaccination at an upcoming on-site clinic (Milkman et al. 2011) and leveraging defaults by introducing opt-out appointment conditions (Chapman et al. 2010, 2016). Combining these tactics with the increasing prevalence of employer sponsored on-site flu clinics shows promise for maintaining high demand for flu vaccinations.

Shifting our attention to the supply-side of the problem, we find that the largest opportunity for improvement, in the giving of shots, is around the "last mile" problem: the administration of the dose. This is largely because many of the other supply issues have already been addressed. For example, perhaps the most memorable shortage of flu vaccinations occurred in the 2004-05 season, when a United Kingdom manufacturer was shut down because of safety compliance issues, resulting in approximately half the required supply being unavailable. Since that shortage, each season the CDC has stockpiled its own supply of flu vaccinations to serve as a safety stock for unexpected occurrences. Additionally, manufacturers now provide the vaccines via rolling production and ship early to provide a near continuous supply to providers. Finally, any motivation to under-order is mitigated with buy-back contracts from the vaccine manufacturers (in line with Pasternack 1985). Although there is an important operations literature discussing improving flu vaccine supply chains (Deo and Corbett 2009, Cho 2010, Arifoğlu et al. 2012), the Food and Drug Administration (FDA) has not noted a significant flu vaccine shortage in the last 5 years (FDA 2019).

Despite the efforts to increase demand and guarantee supply, the flu vaccination rate in the US has remained stagnant in recent years. Adult vaccination rates for flu seasons 2010-11 through 2017-18 were 40.5%, 38.8%, 41.5%, 42.2%, 43.6%, 41.7%, 43.3%, and 37.1%, respectively (CDC 2018). This trend is also unlikely due to patient rejection of provider recommendations. For example, the Affordable Care Act mandated flu vaccinations be covered at zero patient cost for all insurance plans, removing any financial burden. Separately, as Brewer et al. (2017) illustrate, the general rate of vaccine rejection by patients remains stable between 1% and 2% each year, and Patel et al. (2017) find that more than 99% of patients accept the flu vaccine when providers recommend it. Thus, it is unlikely patients are refusing the vaccination when it is offered. The sensible implication is providers simply do not offer as many flu vaccines as they should.

This implication begs the question: "Why are providers not offering more flu vaccinations?" One possible explanation is providers experience severe demands on their time and attention. Prior literature labels physicians who provide traditional patient care as *unfocused*, given not only their lack of specialization, but also the diverse demands on their time (Huckman and Zinner 2008). The truth is these physicians have numerous, competing, demands on their attention. In the 2018 Physician's Foundation survey of more 9,000 physicians, 80% reported over or at-capacity utilization, and 78% reported regularly experiencing feelings of burnout (The Physician's Foundation 2018).

This raises the immediate question: how might one focus provider attention on desired behaviors – for example, flu vaccination? A key element of an operating system is to prioritize tasks of operational decision makers through the design of an operational *structure* which channels attention toward desired tasks and away from peripheral activities (Simon 1947, Hayes & Clark 1988). For example, in an observational (i.e., non-experimental) study, Patel et al. (2017) show that using a "nudge" to cue provider attention by requiring a provider to make a decision (active choice) instead of relying on passive workflow is related to an increase in vaccinations. Thus, the operational *structure* of the organization serves to *focus the attention* of its key decision makers, and this structure can be leveraged through the use of key organizational controls, such as incentives and performance feedback.

The effect financial incentives have on task performance has been well studied in academic literature. Even Taylor (1911) understood the role of incentives in eliciting effort: "…in order to have any hope of obtaining the initiative of his workmen the manager must give some *special incentive*." (p. 14) In behavioral economics, the *law of behavior* states simply, "… [financial] incentives will lead to more effort and higher performance." (Gneezy et al. 2011) This law illustrates a well-known fact: piece-rate incentive schemes are associated with individuals working harder through the adjustment of an agent's perceived utility from an action (Prendergast 1999, Lazear 2000, Shearer 2004). In our specific case, financial incentives seek to align individual healthcare provider behavior with the goals of the clinic.

In addition to altering a provider's utility gains from administering more shots, we propose the provision of performance feedback will also result in focusing providers' attention on flu shots. Prior work highlights that the social context of competition, or the "reference structure," affects participant output (Roels and Su 2014). When examining healthcare providers, Song et al. (2018) show physician productivity increases after publicizing their relative performance (with individual identifiers) presumably because providers allocate attention to the behavior being ranked and with this focus of attention, they alter their actions. Similar, positive effects have also been noted even when relative performance feedback is private (Meeker et al. 2016). Still other literature highlights that introducing performance rankings leads to an increase in effort even when incentives are not tied directly to rankings (e.g., Kuhnen and Tymula 2012, Charness et al. 2014). Because participants care about others' perception of their performance, we expect performance feedback will also serve to focus attention.

Thus, we propose the application of financial remuneration and performance feedback in the context of a well-defined vaccination season will influence and focus providers' attention, resulting in improved performance. Additionally, it is important to note in our context the intervention happens at the *clinic* level and not at the *provider* level. The aforementioned literature focuses on the *individual* response within the context of the *firm* (Patel et al. 2017, Lazear 2000, Shearer 2004, Roels and Su 2014, etc.), whereas little is known about the impact of an such treatments at the firm level. This is another contribution of our work, as we explore an empirical setting with some separation between the intervention and the healthcare provider. One would expect such separation to weaken the overall effect; thus *any* detected effect is a unique contribution.

In our first hypothesis we test incentives and feedback together, relative to a control group, and in the complimentary hypotheses, 1b and 1c, we test each separately relative to a control group.

Hypothesis 1a Intervention clinics who ***simultaneously receive*** financial incentives for reaching growth thresholds and performance feedback relative to peer clinics will show greater Year-over-Year growth in flu shots administered than other clinics in the same state outside the intervention.

Hypothesis 1b Treated clinics who receive financial incentives for reaching growth thresholds will show greater Year-over-Year growth in flu shots administered than the respective control clinics.

Hypothesis 1c Treated clinics who receive performance feedback relative to other clinics in the same treatment group will show greater Year-over-Year growth in flu shots administered than the respective control clinics.

A natural follow-up question to the hypotheses above is which treatment proves most effective in driving growth and performance: financial incentives ***or*** performance feedback? As noted above, there is extensive literature confirming the positive impact of financial incentives (Taylor 1911, Prendergast 1999, Lazear 2000, Shearer 2004, Gneezy et al. 2011). Nevertheless, there are also numerous examples of instances where these incentives can backfire, such as when the introduction of financial incentives "crowd-out" an individual's intrinsic motivation to complete the task of interest (e.g., Fehr and Gächter 2000, Gneezy et al. 2011; Staats et al. 2016). Moreover, additional financial incentives, in cases where individuals feel that their actions are not motivated for financial reasons, could lead to reactance where individuals respond negatively, or at least less positively, than they might otherwise (c.f., Frey 1993). Given that in our case, as clinics are already financially motivated to administer shots, we maintain such a scenario is less likely; our financial incentive is simply a bonus over and above the typical payment. Despite the existence of these counterexamples, the overwhelming conclusion from incentive research is that paying for performance induces increased effort (Prendergast 1999, Lazear 2000, Shearer 2004).

On the other hand, the conclusions from the performance feedback literature are less consistent. As Dechenaux et al. (2015) note in their survey on performance feedback in contests: "The findings on feedback are mixed, with different studies often providing contrasting results." The authors proceed to demonstrate empirical studies with similar settings and conflicting implications. Furthermore, in a setting with a piece-rate incentive scheme similar to ours, Eriksson et al. (2009) find, "…feedback on relative performance, regardless of the performance-pay scheme used, does not improve performance." Moreover, Song et al. (2018) find significant performance improvement when transitioning from private to public feedback, implying that private feedback alone (as in our setting) may not yield sufficient improvement.

The question of which effect will dominate is eventually an empirical one. However, because of the consistent conclusions of improved performance in the incentives literature, and the mixed results from the feedback literature, in our second hypothesis we propose that clinics in a financially incentivized treatment will outperform clinics who are provided with performance feedback relative to other clinics in their respective treatment arm.

Hypothesis 2 Treated clinics who receive financial incentives for reaching growth thresholds will show greater Year-over-Year growth in flu shots administered than clinics who receive relative performance feedback.

In addition to its impact on effort, within the ranking condition we expect distinct behavioral responses to differing ranks. Lower (better) and higher (worse) ranks may lead to different responses to performance feedback, but these specific effects are largely ignored in most literature. One notable exception is Gill et al. (2019), where the authors experimentally characterize a U-shaped *rank response function*. Their participants demonstrated "first-place loving" and "last-place loathing" behaviors: participants ranked near the top and the bottom exerted the most effort, while participants ranked in the middle scaled back their effort. Kuziemko et al. (2014) and Buell (2019) document a similar aversion to worse ranks, specifically focusing on last place.

The literature studying social comparison provides an explanation as to why relative position matters for behavior. Festinger (1954) argues that individuals have an innate desire to evaluate their own capabilities, and the presence of comparative feedback enhances the ability to accurately evaluate one's own performance and progress: social comparison enables better self-evaluation. As a concrete example, one may consider the studies of Home Energy Reports which have demonstrated the efficacy of "descriptive norms" to reduce energy usage by comparing one's own consumption to that of an efficient neighbor (Allcott 2011, Costa and Kahn 2013, Ayres et al. 2013, Allcott and Rogers 2014). Other key conclusions from this line of research are that higher relative position emerges in models of individual decision making (Cole, Mailath, and Postlewaite 1992) and that a rank-response is biologically "hard-wired" (Raleigh et al. 1984). These results extend beyond social-esteem and manifest independently as self-esteem, or the simple "joy of winning" (Coffey and Maloney 2010). And although Kuziemko et al. (2014) mention some of the broader societal concerns of last place, they also highlight the childhood fear of being "picked last in gym class" as perhaps one of the most common occurrences of last place aversion. In any case, the fact remains that relative-rank matters.

To date, the effects of first and last place performance feedback, though significant, have primarily been evaluated in the laboratory. This prior work highlights the need to consider rank response in real contexts with less concrete definitions of first and last place. For example, Kuziemko et al. (2014) question whether their laboratory findings, though supplemented with survey data, will hold in complex business environments since, "…in the real world, the concept of last place is far less well defined than in the… environments described here" (p. 107). Exceptions include Barankay (2011, 2012) and Bandiera et al. (2013) who conduct field experiments but make no effort to study the response to a specific rank. Legge and Schmid (2015) examine professional ski racers in a natural experiment but cannot disentangle the effect of rank from associated prizes. Finally, Buell (2019) answers the call with just such a real-world study, where the author finds customers at the end of the queue at a local supermarket are more likely to abandon. Interestingly, through experimentation, this behavioral effect largely disappears when the individual is no longer at the very end of the line.

Similar to Buell (2019), we consider a real-world setting in our research and further relax the definitions of first and last place, where an aversion to last place or a preference for first place may actually manifest when ranked within some range of the extremes. In so doing, our work is related to Song et al. (2018) who find that when productivity feedback is public (names are known) within an emergency department, individuals at the bottom improve more than others. A key difference in our work is that we examine the dynamic impact of rankings. In some cases, stars or laggards may stay constant. Song et al. (2018) study such a context and so fix their top and bottom performers as a function of the pre-intervention performance. In our context, we are able to study clinics moving in and out of relative rankings and so explore the dynamic effect of relative ranking. Interestingly, we find that more than 20% of the clinics rotate through the bottom 5 ranks, and only 4 of the clinics spend a majority of the vaccination season there. Furthermore, although almost 40% of the clinics move through the top 5 ranks and only 5 clinics spend a majority of the vaccination season in the top 5, one clinic claims the top rank mid-way through the season and never relinquishes that rank. Thus, there is a regular eb and flow in the rankings which provides the opportunity to study the resulting effect on performance. The unique context of our research allows for an empirical examination of specific rank-responses to performance feedback in a real-world setting. In line with prior work, we propose rank-response behaviors will be evident both among high and poor performers. Given these insights, we formalize two complementary hypotheses:

Hypothesis 3 Clinics ranked near the top will increase their vaccination rate to maintain a position amongst high-performing clinics.

Hypothesis 4 Clinics ranked near the bottom will increase their vaccination rate to move away from other poor-performing clinics.

3 Influenza Vaccination & Company Setting

3.1 Vaccination Fundamentals

Vaccination is the process of introducing a weakened, foreign pathogen to the human immune system, provoking a biological response to this pathogen. If effective, the vaccine causes the body's immune system to identify the pathogen as a threat, destroy it and generate future protection (or immunity). This process has two benefits, where the individual being vaccinated is less likely to contract the disease, but also a vaccinated individual is less likely to spread the disease to another. Health professionals recognize this added societal benefit of "herd immunity" when a significant percentage of the population is vaccinated against a particular disease. This immunity serves to protect even members of the community who are not eligible for vaccination. Brewer et al. (2017) note that the vaccination threshold for herd immunity is generally between 80% and 90% but varies by disease.

Most influenza vaccinations, like other vaccinations, are delivered via injection to the upper arm. These injections contain dead influenza microorganisms. As there are numerous strains of flu, the flu vaccine can only protect against 3 or 4 strains each year. In February, the World Health Organization (WHO) convenes flu industry and laboratory experts to recommend which strains should be included for the upcoming vaccination season, which generally starts in the fall. In the US, the Food and Drug Administration (FDA) then decides which strains to include and vaccine manufacturers begin production. A limited number of flu vaccines are initially available to health care providers in August; rolling production continues throughout the fall. As occurrences of influenza like illness (ILI) peak around January or February in the US, it is important to begin vaccinating the population before this time.

3.2 Company Setting

We test our hypotheses with data provided to us by VaxCom (a pseudonym). VaxCom is a vaccine management company headquartered in Florida. They partner with independent health care clinics to manage all vaccination logistics for their customers. This includes ordering vaccines from the manufacturer, managing and tracking clinic inventory, and billing the appropriate payer. VaxCom keeps a percentage of the fee billed to the payer as an administrative fee, and the remainder is passed on to the clinic. Although it is possible a clinic could manage its own process and claim the entire fee, the value proposal lies in the logistic management: most clinics do not rigorously monitor their inventory, which can lead to significant loss. Although this value is particularly attractive for clinics who may be limited by administrative capacity, all clinics stand to benefit from a reduction in vaccine administration hurdles.

For the purposes of this paper, we define flu vaccination season as 1 August through 31 December of a given year. Approximately 95% of VaxCom's flu shots are given within this time window.

3.3 Data Summary & Data Generation Process

VaxCom has provided us with their complete shot record for all their clinics for calendar years 2015, 2016, 2017, and 2018. This data contains information from 2,538 clinics scattered across 16 states and includes over three million observations of administered shots. These records include influenza vaccinations in addition to many others (such as vaccinations for shingles or pneumococcal). The vaccination trends over the years are similar: some vaccinations are given in early August, but vaccinations peak in October and then trend downward as the year concludes.

This data is tracked through VaxCom's technology device, the VaxHub (see Figure 4). When a provider removes a vaccine vial from the vaccine fridge, it is scanned into this device and assigned to the current patient. The provider then administers the vaccine and returns the unused portion of the vial to the fridge. The clinic names and exact locations were de-identified before being provided to us. The shot record has also been stripped of sensitive patient information. See Section 12 (Appendix, Table 9) for a list of information provided for each clinic and the associated vaccination. For reference, the most prevalent shots in this dataset are influenza and there are 53 different influenza vaccinations.

4 Study 1: Description & Empirical Approach

4.1 Study 1: Compliance Care 2017

In August 2017, VaxCom introduced an intervention program aimed at increasing influenza vaccination compliance with the CDC recommendation that everyone over the age of 6 months receive a vaccine each year. VaxCom called this program "Compliance Care" (referred to hereafter as CC). This program paid a rebate of $1, $2, or $3 per shot on top of the above shot payment if a clinic achieved various target thresholds of growth over the previous year. These incentive tiers were structured at set percentage thresholds, targeting Year-over-Year total shot growth; we refer to these thresholds as Tier 1, Tier 2, and Tier 3. The Tier 2 threshold was typically 30% Year-over-Year growth. Because the program targeted improving Year-over-Year growth, bonuses were paid, if earned, at the end of the season.

Compliance Care 2017 was restricted to the state of Florida, and 101 Florida clinics were recruited to participate for the 2017 flu vaccination season while 140 Florida clinics remained outside of CC. Figure 5 illustrates a map with the pinned zip codes of all Florida clinics, and the green pinned clinics indicate clinics enrolled in Compliance Care. Clinics were selected for participation by VaxCom's Compliance Care program director. These clinics were identified based on their history with VaxCom, and if they had proved a reliable partner to date. Clinics were provided access to their progress toward the incentive tier through a web portal; the portal displayed the percent progress to the second-tier incentive. In addition, clinics were also provided with a performance ranking relative to other CC clinics. This ranking was based on the percent progress to the second-tier incentive. Thus, on the web portal, a health care provider could access their own progress and see where they stood relative to other CC clinics in Florida. CC clinics could view their ranking, relative to others, but could not view others' names or performance.

At this time, it is important for us to acknowledge the limitations of Study 1, particularly as it relates to clinic selection potentially driving the positive results. Nevertheless, Study 1 provides valuable insights both to the research team and VaxCom and clearly identifies the need to execute a randomized field experiment. The design of Study 2, and *especially* the clinic consideration criteria, were directly informed by our analysis of Study 1. A complete discussion of the limitations of Study 1 can be found in Section 5.3 and the steps taken in Study 2 to address these concerns can be found in Section 6.1.

4.2 Study 1: Empirical Analysis Strategy

Hypothesis 1a evaluates the initial impact of Compliance Care 2017 and whether the program was effective in increasing clinic vaccination rate for the clinics in the intervention. To begin we simply compared the number of Florida vaccinations from 2016 to 2017, by intervention participation. If the percent increase of vaccinations is greater for the intervention clinics between 2016 and 2017, then the program corresponded to an increase in vaccinations. We follow these descriptive results with a pooled Ordinary Least Squares (OLS) regression model focused on evaluating *cumulative* flu shots each week while clustering the standard errors at the clinic level. In this case, we examine only the clinics within the state of Florida as this controls for the differing demand for flu shots in various regions.

4.3 Study 1: Control Variables

In Study 1, we control for a week fixed effect to account for the weekly trend in demand progression of flu shots; see Figure 2 for the Non-CC weekly flu shot demand in Florida, which appears to peak in October and then taper off into December. Because Study 1 clinics were located exclusively in Florida, we do not control for any other effects.

4.4 Study 1: Independent Variables

In Study 1, in order to address Hypothesis 1a, which focuses exclusively on 2017, we include an indicator for the intervention clinics to measure their difference from other Florida clinics outside the intervention. Note that because all CC clinics were provided with both the incentive thresholds and performance feedback, we cannot evaluate Hypothesis 2 in Study 1. To test Hypotheses 3 and 4, we propose setting binary indicators for clinics ranked among the best or worst performing clinics. As discussed above, with respect to Kuziemko et al. (2014), it is not entirely clear what we should consider “first” or “last” place. VaxCom generated the clinic rankings based on progress to the Tier 2 incentive threshold.[1]

Based on this ranking criterion, we initially propose the most desirable ranks are the top 10 ranks, with “last place” defined as the bottom 10 ranks, or any position in the top or bottom 10%, rounded to the nearest integer. Thus, the clinics ranked in any of the top ten spots that week would be identified as top performers and assigned an indicator (Variable: HighRank set to 1). Similarly, clinics ranked in any of the bottom 10 places would be identified as poor performers and assigned an indicator (Variable: LowRank set to 1). As there is a time delay between a provider checking their ranking and improving their vaccination rate, we lag these variables by one week. These indicators were included in order to test for First-Place Loving (FPL) behavior and Last-Place Aversion (LPA), respectively. As such, we refer to these variables as the FPL/LPA variables.

4.5 Study 1: Dependent Variables and Model Characterization

4.5.1 Cumulative Weekly Shots

The dependent variable to determine the overall treatment effect of Compliance Care 2017 is the cumulative count of flu shots given for each clinic over the course of the season.

$$CumulativeClinicShots_{it} = \alpha_0 + \alpha_1 \mathbb{1}[\,CCclinic\,]_i + Week_t + \epsilon_{it} \tag{1}$$

We evaluated this model for clinic *i* in week *t* using a pooled ordinary least squares (OLS) approach, and the errors were clustered at the clinic level. The **CCclinic** indicator is set to 1 for Compliance Care clinics and a 0 otherwise; this variable (and the resulting coefficient α_1) is the explanatory variable of interest for this model. The **Week** fixed effect controls for any time trend in administering flu shots (see Figure 2 for 2017 trend). The idiosyncratic shocks which we have not controlled for are captured in ϵ.

[1] For example, if a clinic cumulatively administered 80 shots by week 40 in 2017 and the Tier 2 threshold was set at 100 shots, that clinic would have 80% progress and be ranked versus other clinics based on this progress.

4.5.2 Weekly Change in Percent of Clinic Total

In order to test for the presence of rank response behaviors (FPL/LPA) in the CC clinics, we implement an alternate variable which is sensitive to weekly changes. Instead of testing the cumulative effect over the season, we want to know if a CC clinic moves to a high or low rank (defined in Section 4.4) and whether the resulting change in vaccinations is different versus any other week. Within Florida, there is significant heterogeneity in clinic size and capacity to administer shots. In order to allow a comparison of all clinics to one another, we normalize the change in weekly shots by each clinic's total shots given in 2017.

$$PCT\Delta_{it} = \delta_0 + \delta_1 \mathbb{1}[\,HighRank\,]_{i,t-1} + \delta_2 \mathbb{1}[\,LowRank]_{i,t-1} + Week_t + \epsilon_{it} \qquad (2)$$

We evaluated this model using a pooled OLS approach, and the errors were clustered at the clinic level. The outcome variable of interest for clinic *i* in week *t* is the weekly rate (shots per week) change as a percentage of total clinic capacity (PCTΔ).[2] The **HighRank** and **LowRank** indicators are set to 1 if the clinic achieves a ranking in the top 10 or in the bottom 10, respectively, and 0 otherwise. These indicators are lagged by one week to account for the delayed impact of feedback on vaccination behavior. The **Week** fixed effect controls for any time trend in administering flu shots (see Figure 2 for 2017 trend). The idiosyncratic shocks which we have not controlled for are captured in $\boldsymbol{\epsilon}$.

4.6 Study 1: Exclusion Criteria

In total, VaxCom partnered with 244 active clinics in Florida in 2017 and 101 clinics were recruited to participate in Compliance Care, with the remaining 143 clinics outside the program. Of these untreated clinics, 68 clinics were partners with VaxCom through vaccination seasons 2016 and 2017; we refer to these as the Non-CC Year-Over-Year (YOY) partners. We restrict the comparison of the CC clinics to these Non-CC YOY partner clinics, as the remaining clinics possess unknown performance baselines in 2016. For the 101 treated clinics, 23 can be identified as CC YOY partners (full enrollment during the 2016 and 2017 flu vaccination seasons). The remaining 78 CC clinics were not VaxCom partners for the full 2016 season; for these clinics, VaxCom worked with the clinic's administrators and healthcare providers to establish a flu vaccination baseline in 2016 to allow for growth evaluation in 2017. We will present the results for both subsets of CC clinics.

[2] For example, if a clinic gave 20 shots just in week 50 and administered 200 shots in total over the 2018 vaccination season, the weekly change in percent of clinic total would be 20/200 or 10% for week 50.

5 Study 1: Results

5.1 Study 1: Overall Percent Growth & Regression Testing

In Study 1, approximately a third of the clinics reached some level of the incentive threshold. Table 1 details the number of clinics who achieved each incentive threshold in Study 1 (all clinics from CC 2017). See Section 12 (Appendix, Table 11) for flu shot averages by state and by CC participation in 2017 and 2018.

Insert Table 1 here

The flu vaccination totals for Florida clinics in 2016 and 2017 are illustrated in Table 2. There are 68 clinics in Florida outside of Compliance Care for which we have full shot information for flu vaccination seasons 2016 and 2017. The last CC row presents the Year-Over-Year (YOY) partners for which we have complete flu vaccination information for 2016 and 2017. This list of 23 clinics is the most robust as well as the most conservative grouping of CC clinics.

Insert Table 2 here

Overall, Table 2 demonstrates that outside of Compliance Care, Florida administered fewer flu shots in 2017. We propose this is due, at least in part, to Hurricane Irma making landfall in Florida on September 10, 2017. We observe this disruption in the data as some clinics in affected areas experience dramatic drops in shots given which also correlates with reports of significant power outages. To understand the impact of CC we examine the net effect, the difference in percent growth between each CC subgroup and the rest of the state. In the most conservative estimate, the YOY Partner CC clinics were approximately flat, for a net effect of +8.94%.

In order to demonstrate further robustness, we evaluated a pooled OLS regression measuring the cumulative change in clinic doses over 2017 while controlling for weekly fixed effects. The end result versus the Non-CC clinics in Florida was the following: + 61.8 shots (or + 22.5%) when comparing all CC clinics to Non-CC YOY partners ($p = 0.003$); + 77.1 shots (or + 25.0%) when comparing YOY CC partners to Non-CC YOY partners ($p = 0.049$). On average, this indicates CC clinics cumulatively administered about 20% more shots more than Non-CC clinics over 2017, and this result is statistically significant. All of these results imply Compliance Care 2017 had a positive net effect which was related to an increase in the number of flu shots administered in Florida. This provides support for Hypothesis 1a, without accounting for selection bias.

5.2 Study 1: Rank Response

Given the above definition for a high rank (top 10 ranks) and low rank (bottom 10 ranks), we find that moving into these positions is followed by an increase in vaccinations, measured as the weekly change in the percent of clinic total shots. When evaluating FPL behavior for all 101 CC clinics, a clinic which achieves a ***high*** rank experiences a +0.08% change in shots given (all results relative to clinic capacity) the following week, but the results are not statistically significant (p = 0.745). Similarly, when evaluating FPL for the 23 YOY partners, a clinic which achieves a ***high*** rank experiences a *negative* change in shots given the following week. When evaluating LPA for all 101 CC clinics, a clinic which achieves a ***low*** rank experiences a +1.20% change in shots given and the results are statistically significant with a p-value of 0.084. Similarly, when evaluating LPA for the 23 YOY partners, a clinic which achieves a ***low*** rank experiences a +2.18% change in shots given the following week, and the results are statistically significant with a p-value of 0.004. We test the sensitivity of our findings to these definitions of first and last place in Section 13, with similar results.

According to the above definitions of first and last place in addition to the sensitivity testing of these definitions, these results appear to indicate the presence of Last-Place Aversion. Unfortunately, we cannot say there is sufficient evidence of First-Place Loving behavior, as the clinics do not exhibit FPL behavior as we have defined here. Thus, in Study 1, we find no evidence to support Hypothesis 3 and strong evidence to support Hypothesis 4. We proceed to test the robustness of these findings with the randomized treatment in Study 2.

5.3 Study 1: Limitations

The key limitation of Study 1 is that clinics were not assigned randomly to the treatment and control groups. The goal of CC's program director, as explained to us, was to find clinics that would take part in the intervention. Given this screening criteria there is a meaningful risk that those clinics more likely to seek growth might then take part in the intervention. However, we note that the test of Hypotheses 3 and 4 is not subject to this selection bias concern as we look only within CC treatment clinics.

Other limitations arise as a factor of the implementation of the program and the information availability of participants. For example, several clinics were added as VaxCom partners late in the 2016 vaccination season; thus, in order to establish a baseline for growth in 2017, this required the estimation of their vaccination history. As a result of this less reliable baseline, the Year-over-Year percent growth goals were not uniform across all CC clinics. Most clinics (43 of 101) were given a 30% goal for the Tier

2 threshold, however some clinics with limited flu vaccine history were given more modest Tier 2 goals (15% Year-over-Year growth) while other clinics were given "stretch" goals (more than 50% growth) to encourage these clinics to identify their true capacity to administer shots. Because of the heterogeneity among clinic performance baselines and incentive thresholds, robustly evaluating performance becomes difficult. Finally, the simultaneous implementation of both feedback and incentives confounds our ability to identify the distinct effect of each intervention element. Despite these limitations, Study 1 provides an initial exploration of the impact of financial incentives and rankings on vaccinations. The study also provides important inputs for the field experiment, including randomization, conducted in Study 2.

6 Study 2: Description & Empirical Approach

6.1 Study 2: Compliance Care 2018

Informed by the outcomes of Study 1, Compliance Care 2018 ("Study 2") was launched in August 2018. Study 2 included 145 clinics spread across 9 different states. Within each state, participant clinics were identified and then randomly selected into three approximately equally sized treatment arms: Rebate (49 clinics), Ranking (47 clinics), and Control (49 clinics). To accomplish this task, we took the following steps.

First, in order to determine the appropriate sample size in Study 2, we performed a power analysis on the difference of means between the CC and Non-CC clinics in Florida in Study 1. By setting acceptable rates for Type I and Type II errors [(α, β) respectively set at (0.05, 0.2)], our power analysis required a sample size of at least 42 control clinics and 44 clinics in each treatment arm. We calculated this value assuming a case where the control clinics exceed the performance of the treated clinics to be equivalent to "no effect of the treatment" using a one-tailed t-test. Given these results, we rounded these counts up to 50 clinics in each group (150 clinics total) to ensure we have sufficient power to detect an effect.

Second, the researchers worked closely with VaxCom to implement the process of randomly choosing which clinics would be selected for Study 2. The baseline criteria for consideration to participate in CC 2018 were the following: the clinic designation was either Family Practice or Internal Medicine to provide standardization across clinic types; the clinic did not participate in CC 2017 to provide a clean estimate of the treatment effect; the clinic was enrolled for the full 2017 Flu Season to provide an accurate baseline; and the clinic administered at least 150 flu shots during the 2017 Flu Season to verify clinic ability to administer a minimum number of shots in 2018 and so participate meaningfully.

Third, with the consideration criteria determined VaxCom next selected 9 states to serve as strata within which clinics could randomly be assigned to treatment. A target number of clinics was selected for each state based on the state's overall size. The leadership team then solicited the state-level salesforce for recommendations on which clinics would be representative participants in order to meet the consideration criteria. This helped filter out clinics that were unresponsive partners or behind on payments to VaxCom. This group of clinics was randomly sorted into treatment groups at the state-level. The state-level representatives then solicited each clinic's participation in CC 2018 for each clinic's designated treatment arm. Though participation *required* no additional cost on the part of the clinic, consent was still necessary. With clinic agreement, CC 2018 commenced in August 2018; please see Section 12 for a group comparison of pre- and post-Compliance Care 2018 (Appendix, Table 10). Other Appendix tables illustrate the breakdown of clinic participants by treatment group by state (Table 12, Table 13, Table 14). These tables include clinic participation after recruitment, at the beginning of the vaccination season, and at the end of the season. Before the beginning of the season, 5 clinics terminated their partnership with VaxCom. In total, 10 clinics terminated before the end of the 2018 vaccination season. Clinics have been removed from the analyses where indicated.

After recruitment, no further intervention was taken with the Control clinics. The Rebate clinics received tiered incentive thresholds similar to those described in Study 1, with rebates of $1, $2, and $3 per shot awarded for achieving the targets set at 10%, 15%, and 20% Year-over-Year growth across all clinics. The Ranking clinics were provided with weekly rankings based on their overall Year-over-Year percentage growth versus the other clinics in the Ranking treatment. Unlike in Study 1, the treated clinics were sent a weekly email which contained their relevant information: ranking for the Ranking clinics and percent growth and incentive thresholds (both the percent growth required and the associated shot count) for the Rebate clinics. Please see Figure 6 and Figure 7 for reporting templates for the Rebate and Ranking group, respectively.

Given this approach, Study 2 addresses all the major concerns that were present with Study 1. First, strict selection criteria were implemented to establish a uniform assortment of clinics. This removed the requirement to estimate vaccination history and allowed for robust measures of clinic growth. Second, geographically dispersed clinics were randomly sorted into each treatment arm, mitigating the concern of selection bias due to either performance history or location. Third, random selection allowed for the creation of a distinct, comparable control group, instead of using screened-out clinics as a proxy for a control. Fourth, the treatments (financial incentives and performance feedback) were separated into distinct groups to allow independent evaluation against the control. Fifth, in the Rebate group, all of the clinics received identical incentive thresholds (10%/15%/20% Year-over-Year growth). Finally, the use of email notifications removed the requirement for clinic staff to exert additional effort to check their vaccination progress via an online portal. Thus, Study 2 permits a more robust evaluation of each treatment to determine the effect of the interventions.

6.2 Study 2: Empirical Analysis Strategy

Hypothesis 1b and 1c address whether the Rebate and Ranking groups, respectively, outperformed the Control group in the randomized experiment. We initially explore the differences in the summary statistics, by condition and then we perform further longitudinal regression analysis to rigorously evaluate the results and control for additional elements. All errors were clustered at the clinic level. Given the results from this model, we can also evaluate Hypothesis 2 and test whether the Rebate treatment arm outperformed the Ranking arm.

For our final hypotheses (H3 and H4), we evaluate the behavioral patterns within treated clinics. Following the structure of our panel data, we utilized a fixed-effect regression model, focusing on *weekly changes* in the dependent variables instead of cumulative changes. We are specifically interested in First-Place Loving and Last-Place Aversion behaviors, and we construct independent variables to test for the presence of these behavioral patterns.

We include the variables described below to test our hypotheses.

6.3 Study 2: Control Variables

When evaluating Hypothesis 1b, Hypothesis 1c, and Hypothesis 2, because the clinics and treatments are dispersed geographically across 9 states, we control for clinic fixed effects in a Fixed Effects panel regression model and through direct estimation in a Random Effects panel regression model. Additionally, we also control for State fixed effects in a Random Effects panel model, though these effects are dropped in a Fixed Effects model due to their time invariant nature relative to the clinic fixed effect. When evaluating Hypothesis 3 and Hypothesis 4, we control for a weekly fixed effect to account for the weekly trend in demand progression of flu shots; see Figure 3 for weekly flu shot demand over the season, which appears to peak in October and then taper off into December.

6.4 Study 2: Independent Variables

In Study 2, because of the implementation of a randomized experiment before the start of the 2018 vaccination season, the independent variables of interest are straightforward: we measure the difference between the treatment and control groups in 2018. This can be measured by interacting a multi-leveled factor variable for treatment with an indicator for year. The resulting output will measure the difference between the groups by year and serve to evaluate Hypothesis 1b and Hypothesis 2.

To test Hypotheses 3 and 4, we set binary indicators for clinics ranked among the best or worst performing clinics. As before, it is not entirely clear what we should consider "first" or "last" place. In this case, we propose a threshold and test that standard with sensitivity analyses. Different from Study 1, VaxCom generated the clinic rankings based on percent growth in 2018 versus the same week in 2017.[3] Based on this ranking criterion and to align with the analysis of Study 1 with a reduced sample size, we initially define “first place” as the top 5 ranks, with “last place” defined as the bottom 5 positions. This definition aligns with the definitions of Study 1, which initially focused on the top and the bottom 10% of rankings, rounded to the nearest integer. As mentioned above, more than 20% of the clinics rotate through the bottom 5 ranks, and only 4 of the clinics spend a majority of the vaccination season there. For the high performers, almost 40% of the clinics move through the top 5 ranks and only 5 clinics spend a majority of the vaccination season in the top 5, though one clinic achieves this rank mid-season and retains that rank through the end of the season. The clinic ranked in the top 5 ranks each week would be identified as a top performer and assigned an indicator (Variable: HighRank set to 1). Similarly, clinics ranked in any of the last 5 places would be identified as poor performers and assigned an indicator (Variable: LowRank set to 1). As the email arrives at the end of the week and improvement then would follow, if any, the next week, we lag these variables by one week. These indicators were included in order to test for First-Place Loving (FPL) behavior and Last-Place Aversion (LPA), respectively.

6.5 Study 2: Dependent Variables and Model Characterization

Our primary metric of interest for Study 2 is the difference in clinic vaccinations between the treatment and control groups from 2017 (before the introduction of CC 2018) to 2018. We want to know if clinics are increasing the number of vaccinations given each season in response to the feedback given through the Compliance Care program. This rate (shots per season) change can be observed at any time interval. To imitate the weekly update emails provided by VaxCom, we have aggregated the data and observed the rate change by week. This period is long enough to make sure that clinics can examine the feedback (as reported by VaxCom), but still short enough to capture intertemporal differences.

6.5.1 Shots per Patient-Population (DV1)

Despite the implementation of random selection, there are several factors to be mindful of when comparing clinics. One primary concern is that some clinics could be experiencing disproportionate changes in their patient population demographics. For example, a clinic in a remote area of Ohio with a slowing economy might experience greater patient attrition than a clinic in Florida with a more stable patient population. A second concern is clinic capacity, as the CC clinics vary significantly in size and

[3] For example, if a clinic cumulatively administered 110 shots by week 40 in 2018 and 100 cumulative shots by week 40 in 2017, their respective growth would be 10% in week 40 2018 and they would be ranked against other clinics using this metric.

their ability to deliver flu shots. Some clinics are capable of administering 2,000 flu shots over the flu season, while others might struggle to deliver 200 shots.

In order to account for both these factors and to allow for appropriate comparison of these clinics to one another, we divide the cumulative number of weekly flu shots given by each clinics' total patient population (Shots per Patient-Population, or SPP) over the vaccination season. This patient count is the count of all patients who sought any service from the clinic, not just patients who received vaccinations. Using this as our primary outcome of interest has two primary advantages. First, this metric controls for changes in patient population between 2017 and 2018. Thus, in the case of patient attrition between 2017 and 2018, 200 shots given while seeing 100 patients is comparable to 100 shots given with 50 patients. The reverse is also true in the case of patient growth: 200 shots given while seeing 100 patients is comparable to 300 shots given with 150 patients. The second advantage of this metric is all clinics' shots are normalized by patient volume, allowing comparison of both large and small clinics.

$$ShotsPerPatientPop_{it} = \beta_0 + \beta_1\, TreatmentGroup_i\, X\, Year_t + Year_t + \epsilon_{it} \qquad \text{(3a, FE)}$$

$$\begin{aligned} ShotsPerPatientPop_{it} = \beta_0 + \beta_1\, TreatmentGroup_i\, X\, Year_t + Year_t \\ + ClinicStateFE_i + ClinicFE_i + \epsilon_{it} \end{aligned} \qquad \text{(3b, RE)}$$

We present the resulting Shots per Patient-Population for clinic *i* in week *t* using a Fixed-Effects model (FE-DV1) and a Random-Effects model (RE-DV1) to allow for additional controls and to evaluate the robustness of Model FE. The **TreatmentGroup** variable is a multi-leveled factor variable with a level for each treatment (Control/Ranking/Rebate). The **Year** variable controls for the differing time trend between 2017 and 2018. When this variable is interacted with the **TreatmentGroup** variable, the resulting coefficient specifies the performance difference between each treatment and the control between 2017 and 2018. In the Random-Effects model, we can also include state fixed effects (**ClinicStateFE**) and clinic fixed effects (**ClinicFE**) to control for the differing demand patterns faced by each clinic and each geographic region.

6.5.2 Percent Growth (DV2)

While the SPP metric presents several advantages, VaxCom is not simply interested in increasing a clinic's shots given per patient: their desire is to drive overall clinic growth. Thus, simply measuring a change in shots is slightly unsatisfactory. As a robustness check and to augment the above metric to measure clinic growth, we also calculate the weekly percent growth of clinics over the 2018 vaccination season. The baseline for this growth is the total number of shots given in 2017. The cumulative weekly doses given is then represented as percent growth from this point.[4] Because this variable is relative to a clinic's own performance in 2017, this variable also helps to control for the differences in clinic size and capacity. To allow for changing patient volume, we can still include a control for patient count in this model. In this case, because we are focused on growth only in 2018, we do not control for a yearly demand trend.

$$\begin{aligned} PercentGrowth_{it} = \beta_0 \ + \ \beta_1 \, TreatmentGroup_i \, X \, Year_t \\ + ClinicState_i + PatientCount_i \ + \ \epsilon_{it} \end{aligned} \tag{4}$$

We present the resulting Percent Growth for clinic *i* in week *t* using a Random-Effects model (RE-DV2). The **TreatmentGroup** variable is a multi-leveled factor variable with a level for each treatment (Control/Ranking/Rebate). The **Year** variable controls for the differing time trend between 2017 and 2018. When this variable is interacted with the **TreatmentGroup** variable, the resulting coefficient specifies the performance difference between each treatment and the control between 2017 and 2018. **ClinicState** controls for geographic differences in vaccination demand, and **PatientCount** controls for each clinic's total patients seen during the vaccination season.

6.5.3 Weekly Change in Shots per Patient-Population (DV3)

In order to test for the presence of rank response behaviors (FPL/LPA) in the Ranking clinics, we implement an alternate variable which is sensitive to weekly changes. Instead of testing the cumulative difference between the treatment and control over the season, we want to know if a clinic in the Ranking treatment moves, in response, to a high or low rank (defined in Section 6.4) and whether the resulting change in vaccinations is different versus any other week. As mentioned previously, there is significant heterogeneity in clinic size and capacity to administer shots. In order to allow a comparison of these clinics to one another, we implement a variant of DV1 (shots per patient-population) which is sensitive to weekly changes and patient volume. Instead of testing the cumulative difference between the treatment and control over the season, we want to know if a clinic moves to a high or low rank (defined in Section 6.4), whether the resulting change in shots per patient-population (SPPΔ) is different versus any other week. This variable is constructed by taking the count of shots given just that week and dividing by the patient count during the associated vaccination season.[5]

$$SPP\Delta_{it} \ = \ \pi_0 \ + \ \pi_1 \, \mathbb{1}[\, HighRank \,]_{i,t-1} \ + \ \pi_2 \, \mathbb{1}[\, LowRank]_{i,t-1} \ + \ Week_t \ + \ \epsilon_{it} \tag{5}$$

[4] For example, if 165 shots had been given up through the end of November 2018 and 150 total shots were given in 2017, this variable would list as 10% growth during that week.
[5] For example, if a clinic gave 20 shots just in week 50 saw 200 patients in total over the vaccination season, the weekly change in shots per patient-population would be 20/200 or 0.1 during week 50.

We evaluate this specification using a Fixed-Effects model, and the errors were clustered at the clinic level. The outcome variable of interest is the weekly rate (shots per week) change as a ratio of total clinic patient-population (shots per patient-population per week, SPPΔ). The **HighRank** and **LowRank** indicators are set to 1 if the clinic achieves a ranking in the 5 ranks or in the bottom 5 ranks, respectively, and 0 otherwise. These indicators are lagged by one week to account for the delayed impact of feedback on vaccination behavior. The **Week** fixed effect controls for any time trend in administering flu shots (see Figure 3 for 2018 trend). The idiosyncratic shocks which we have not controlled for are captured in ϵ.

6.5.4 Weekly Change in Percent of Clinic Total (DV4)

As an alternate test for the presence of rank response behaviors (FPL/LPA) in the Ranking clinics, we normalize the change in weekly shots by each clinic's total shots given in 2018; this gives a weekly rate (shots per week) change as a percentage of total clinic capacity.[6]

$$PCT\Delta_{it} = \pi_0 + \pi_1 \mathbb{1}[\, HighRank \,]_{i,t-1} + \pi_2 \mathbb{1}[\, LowRank]_{i,t-1} + Week_t + \epsilon_{it} \qquad (6)$$

We evaluate this specification using a Fixed-Effects model, and the errors were clustered at the clinic level. The outcome variable of interest for clinic *i* in week *t* is the weekly rate (shots per week) change as a percentage of total clinic capacity (PCTΔ). The **HighRank** and **LowRank** indicators are set to 1 if the clinic achieves a ranking in the top 5 ranks or in the bottom 5 ranks, respectively, and 0 otherwise. These indicators are lagged by one week to account for the delayed impact of feedback on vaccination behavior. The **Week** fixed effect controls for any time trend in administering flu shots (see Figure 3 for 2018 trend). The idiosyncratic shocks which we have not controlled for are captured in ϵ.

6.6 Study 2: Exclusion Criteria

Despite the strict consideration criteria for Study 2 and the benefits derived from participating in Compliance Care 2018, several clinics decided to end their partnership with VaxCom during Study 2. In total, 5 clinics terminated their relationship before the flu vaccination season began in August and another 5 clinics terminated before the end of the season. In total, six clinics terminated from the Control group, while two clinics each terminated from the Ranking and Rebate arms. See Section 14 (Appendix), Table 12, Table 13, and Table 14 for the clinic distributions by group and state over the season.

[6] For example, if a clinic gave 20 shots just in week 50 and administered 200 shots in total over the 2018 vaccination season, the weekly change in percent of clinic total would be 20/200 or 10% for week 50.

These terminations reduced our clinic sample size from 145 to 135; since our initial sample was not large, the loss and exclusion of these 10 clinics will have a negative impact on the statistical power of our results. Given that random selection was performed for all clinics in Study 2 and that all groups suffered from some attrition, retaining the clinics who terminated after the start of the season under the "intent to treat" paradigm is consistent with a typical econometric approach to treatment effects. As researchers, we can only control the setup and application of the treatment and not whether a participant complies with the program. Our primary results will be presented while including all clinics who started the vaccination season as VaxCom partners; this data contains vaccination information for 140 clinics. We thus exclude the clinics who enrolled in Compliance Care 2018 but failed to participate while still including the clinics who participated for a portion of the time period. In total, we refer to these three different clinics subsets: 1) the "Full" clinic set (145 clinics), 2) the "Partial Exclusion" subset (140 clinics), and 3) the "Exclusive" subset (135 clinics). For additional robustness, we demonstrate the outcomes of our models across all three subsets.

7 Study 2: Results

7.1 Study 2: Overall Percent Growth – Descriptive Statistics

In Study 2, similar to Study 1, approximately a third of the clinics reached at least the lowest level of the incentive threshold. Table 1 details the number of clinics who achieved each incentive threshold in Study 2 (Rebate clinics only from CC 2018).

The overall results are shown in Table 3, with the total shots given by clinics in these categories in 2017 (before the start of Study 2) and 2018 during the study. As above, the percent growth columns illustrate how much each group changed year over year and the net effect is the difference between each of the treatment arms and the control. Although both treatment groups showed positive growth, the Ranking treatment shows the most noticeable difference from the Control group. We observe from this table that the Ranking group noticeably outperformed both the Control group and the Rebate treatment.

Insert Table 3 here

However, as discussed in Section 6.6, 5 clinics terminated their partnership with VaxCom before the start of the 2018 vaccination season. By removing these clinics from consideration, one can see the Ranking group continues to exhibit the greatest growth, though with a smaller margin over the Control group, which also experienced positive growth (Table 4). We note the following key points from this table. First, the Rebate treatment group underperformed the Ranking group, raising questions about H2.

Secondly, the Rebate treatment group barely outperformed the Control group. It is reasonable to question whether there is a statistically significant difference between this group and the control. Overall, the summary statistics suggest that, at least, the Rebate treatment may have led to an increase in flu shots, however, it is important to use a robust statistical model to evaluate statistical significance.

Insert Table 4 here

7.2 Study 2: Regression Analysis – Shots per Patient-Population (DV1)

We first examine the impact of our treatments on a clinic's cumulative SPP captured each week using the partial exclusion sample with 140 clinics. Table 5, under the SPP (Shots per Patient-Population) outcome variable, shows the results of both a Fixed Effects (FE-DV1) and Random Effects (RE-DV1) model.

Insert Table 5 here

Table 5 reveals that there is a statistically significant difference between the Ranking and Control clinics when evaluating SPP for DV1 using a Fixed Effects ($p = 0.0495$) and a Random Effects ($p = 0.0503$) model. These results show, on average, the Ranking clinics administered approximately 5% more of their respective patient population than the Control clinics while the Rebate clinics show no meaningful difference. Thus, we see support for Hypothesis 1c, but not Hypothesis 1b. Recall this patient count is the clinic's total patient count; as such, it must be less than or equal to the count of patients who received flu shots, requiring an outcome less than one.

7.3 Robustness Check: Shots Per Patient-Population

Although the results from the previous Section indicate clinic improvement, we now address the robustness of our findings. We evaluate the robustness of our clinic selection criteria given the SPP outcome variable (DV1). As mentioned in Section 6.6, 145 clinics were initially recruited; of these clinics, 5 terminated before the start of the vaccination season and another 5 terminated before the end of the season. Thus, two alternative clinic sets exist to further test the above models: a clinic set which includes all clinics regardless of program status (145 clinics, or the "Full" set) and a clinic set which excludes any clinics which terminated at any point before the end of the vaccination season (135 clinics, or the "Exclusive" set). Please see Table 6 for the results.

Insert Table 6 here

When examining the Full clinic set, our results are largely confirmed. The coefficients for the Ranking group are of slightly greater magnitude (6.55% vs 4.58%) and exhibit greater statistical significance ($p = 0.012$). This is possibly due to the greater sample size we have in this case. Our above conclusions are thus robust to the case when including all clinics: Compliance Care 2018 led to an increase in flu shots for the Ranking group. Turning to the results for the Exclusive clinic set we find that the coefficients of interest are slightly smaller – 3.59% vs. 4.58%. In addition, the p-values are now 0.104 versus 0.049, indicating only marginal significance. Although we acknowledge the research community's general acceptance of the 0.05 threshold for p-values, we would note this threshold is still a subjective norm, and one would be unwise to discard results which fall on one side while unconditionally accepting results on the other side. As such, we recognize these values as continuous measures (Cortina and Landis 2011) and, per Simmons et al. (2011, 2012), we fully disclose our process and results (including any limitations) accordingly. Thus, we see continued general support for our hypothesis that Compliance Care 2018 led to an increase in flu shots for the Ranking subgroup, although the smaller sample size weakens that support. In terms of attrition, we note and emphasize that all groups experienced some clinic attrition. It would be more concerning if the terminations were concentrated in only one group.

7.4 Study 2: Regression Analysis - Percent Growth (DV2)

An alternative way to conceptualize the outcome variable is with the total number of flu shots. This is an important metric for VaxCom in driving both public health and their business growth. As such, we extend our analysis to ascertain if treated Compliance Care clinics exhibited meaningful, percentage growth between 2017 and 2018. We evaluate the difference between the treatment and control groups as measured by change in cumulative percent growth (RE-DV2) for all clinics who started the vaccination season as VaxCom partners. Please see Table 7 for the results.

Insert Table 7 here

Similar to above, there is a meaningful statistical difference between the Ranking and Control clinics when evaluating Year-over-Year percent growth. These results show, on average, the Ranking clinics grew by approximately 8% more than the control clinics ($p = 0.045$) while the Rebate clinics show no meaningful difference in growth from the control. These results are in-line with the percentages shown in Table 4 for the Ranking group (just under 10% growth).

7.5 Comparison between Ranking & Rebate Groups

From the above analyses, the statistical separation is evident between the Ranking and Control groups: the Ranking group clearly outperformed the Control group, demonstrating the significant impact of performance feedback in driving flu shot growth. The more surprising result is the Ranking group, which only received performance feedback relative to their peers, appears to have also outperformed the incented Rebate group. One would not only expect the Rebate group to outperform the control (not evident above) but also to outperform the Ranking group. In the above analysis, this does not appear to be the case and raises the question of whether the difference between the treatment groups is statistically significant.

We can test the coefficients from these regressions in two separate manners. The first option is to perform a Wald test on the coefficients under the null hypothesis that the Rebate performance is greater than or equal to the Ranking performance in CC 2018 (see Hypothesis 2). Alternatively, we can test the linear combination of the coefficients ($\beta_{1,rebate}$ - $\beta_{1,ranking}$) and construct a confidence interval for this difference (Stata: 'lincom'). In this case, a large, positive confidence interval would indicate the Rebate group indeed outperformed the Ranking group, whereas a large, negative interval would indicate exactly the opposite, that the Ranking group outperformed the Rebate group.

We perform both tests on FE-DV1 (shots per patient-population) and RE-DV2 (percent growth). For FE-DV1, under the Wald test we reject the null that the Rebate coefficient is greater than or equal to the Ranking coefficient (p = 0.044). Similarly, for RE-DV2, under the Wald test we reject the null that the Rebate coefficient is greater than or equal to the Ranking coefficient (p = 0.025). The conclusions surrounding the test of the linear combination of coefficients are the same as the Wald test.

Given the results of these tests, we can conclude not only that the Rebate group did *not* statistically outperform the Ranking group but it actually appears the opposite is true: the Ranking group outperformed the Rebate group, and the difference is statistically significant. Thus, to our surprise, we are able to reject Hypothesis 2 at $p < 0.05$ in all cases.

7.6 Study 2: Rank Response Behavior

We next evaluate the presence of rank response behaviors within the Ranking treatment group.

7.6.1 Rank Response Behavior

Given the above definition for a high rank (top 5 ranks) and low rank (bottom 5 ranks), we find that moving into these positions is followed by an increase in vaccinations for DV3 (weekly change in SPP) and DV4 (weekly shot change as percent of clinic total).

When evaluating DV3, a clinic which achieves a high rank leads to an increase of +0.46% change in shots per patient population the following week, but the results are not statistically significant ($p = 0.425$). However, a clinic which achieves a low rank leads to a +2.75% change in shots per patient population the following week, and the results are significant ($p = 0.0009$). When evaluating DV4, a clinic which achieves a high rank, leads to an increase of 0.38% change in shots given (relative to clinic capacity) the following week, but the results are not statistically significant ($p = 0.677$). However, a clinic which achieves a low rank leads to a +3.96% change in shots given (relative to clinic capacity) the following week, and the results are significant ($p = 0.004$). According to the above definitions of first and last place, these results appear to indicate the presence of Last-Place Aversion but do not indicate the presence of First-Place Loving behavior as we have defined it here.

7.6.2 Rank Response: Sensitivity Analysis

We next consider the sensitivity of our conclusions to our definitions of first and last place. To evaluate our thresholds, we fix one parameter at the original definition and vary the other parameter. For example, we continue to reserve the high rank flag for clinics ranked in the top 10 but then relax and restrict the threshold for the bottom ranked clinics to see if there is Last-Place Aversion also amongst clinics who are ranked in the bottom 10. We do this both for DV3 and DV4. Although there is a lack of evidence for FPL behavior in the above analysis (Section 7.6.1), we can further explore the possibility of this behavior here.

We first discuss the results of iterating these thresholds with respect to the definition of first place. We consider any threshold for first place to include restricting only to the top rank and relaxing to being ranked in the top 10 for both dependent variables. The results are statistically significant when first place is defined as either of the top two ranks when evaluating DV3 ($p = 0.005$ and $p = 0.098$) and as the top rank *only* for DV4 ($p = 0.0248$); in all other definitions, the results are not statistically significant. Thus, it is possible if we restrict our definition of first place, we may have evidence of FPL behavior. We further evaluate the robustness of this finding in Section 7.6.3.

The results for iterating the definition of last place can be seen in Table 8. Note that the results for last place presented in Section 7.6.1 can be seen in column 7. When evaluating our definition of last place, we evaluated the cutoff down to clinics that were ranked in the bottom half. From Table 8, we still see evidence for LPA when restricted to only considering the clinic which ranked last. Additionally, evidence exists for LPA even when relaxing the definition to consider clinics ranked in the bottom 10. Both dependent variables flag a difference in the clinic's response when defining "last place" around the bottom 10 ranks. It is important to observe that the definition of last place does not continue through all possible relaxations; were this the case, one could make the argument that we are not measuring our intended parameter, as one would not expect LPA to exist in, for example, the clinics ranked closer to the top.

Throughout this analysis, we maintain the robustness of our last place definitions and their results. Using two separate outcome variables, we find evidence of Last-Place Aversion amongst the clinics who received relative performance feedback in Study 2 and the possible presence of First-Place Loving behavior.

7.6.3 Rank Response: Robustness

Because of the straightforward application of indicator variables to test for the presence of these behaviors, there could be alternate mechanisms behind the variance we explain with these variables. One important alternative to consider is mean reversion; clinics are likely to have weeks with poor performance followed by high performance, either because of the fluctuation in demand or a provider's attempt to "catch back up." If this behavior coincided with changes in rank, it is possible our explanatory variables could be picking up this reversion and that is not simply FPL or LPA.

To test for this, we artificially ranked the clinics in the Control group (with respect to each other) and applied the same models (FE-DV3 and FE-DV4) to these clinics, similar to a Placebo test. Similar to the sensitivity analysis of Section 7.6.2, we test for all definitions of the first place to clinics ranked in the top 10. When testing these definitions in the Control clinics, we do see instances of statistically significant results when using both DV3 and DV4 in the top one or two ranks. When evaluating DV3, a control clinic which achieves the top two (one) ranks leads to a of +1.04% (+1.20%) change in SPP the following week, and the results are statistically significant with a p-value of 0.002 (0.039).When evaluating DV4, after a control clinic achieves the top one rank, it leads to an increase of 1.75% change in shots given (relative to clinic capacity) the following week, and the results are significant with a p-value of 0.051.

When considering LPA, similar to the sensitivity analysis of Section 7.6.2, we test for LPA under any definition of the last place cutoff down to clinics that were ranked in the bottom half (ranks greater than or equal to 23 out of 46 Control clinics). When testing for LPA, we do not see any statistically significant results among the Control clinics for any of these definitions. Additionally, the coefficient values for the Control clinics appear to be negative when using either DV (though not significant).

Given these results, we do conclude that there is sufficient evidence for Last-Place Aversion, as the behavior we measure among the treated Ranking clinics does not manifest in the Control clinics. However, these results do indicate that we are not explaining the performance of the clinics who achieve a high rank, as there is some behavior for these top ranked clinics which is common between the Ranking and Control clinics, who received no performance feedback during the season. One explanation for why we do not see FPL behavior is perhaps the top ranked clinics have vaccinated the maximum number of possible patients in their area of influence. When we examine the top 10 ranked clinics at the end of the season, 5 of these clinics exhibit conversion rates (defined as total shots given divided by total patient population) over 50% and of these, 3 clinics exhibit greater than 80% conversion rates. It is possible these high performing clinics have saturated their available market.

Thus, in summary, we do not find enough to robustly claim evidence for Hypothesis 3 and First-Place Loving behavior, *however,* we do find strong evidence to support Hypothesis 4 and the presence of Last-Place Aversion.

7.7 Overall Implications of Relative Performance Feedback

We have demonstrated that the introduction of Compliance Care 2018 did in fact lead to an increase in flu shots, most notably for the group which received relative performance feedback. As the vaccination season progressed, the clinics competed with one another to achieve a higher rank or perhaps simply to avoid falling too far behind. But the question remains regarding the *manifestation* of improvement. We know the Ranking treatment led to improvement. But within this treatment, exactly which clinics improved? Was it the clinics who were already high performers and immediately set out to achieve a high rank early? Or by introducing the feedback, have we pulled up the performance of the clinics who would generally have watched the season progress without exerting extra effort?

If we use the artificial rankings for the Control clinics (Section 7.6.3) and match these to the Ranking clinics based on their respective rankings at the end of the season, we can observe the location of the most improvement. Please see Figure 8 for reference. With two exceptions (in Ranks #1 and #34), the Ranking clinics outperform their similarly ranked Control clinics. Across the board, the Ranking clinics outperform their respective Control clinics by 10.3%. For the Ranking clinics with negative performance (administered fewer shots in 2018 versus 2017), the Ranking clinics outperform their respective Control clinics by 22.9%. For the Ranking clinics who showed positive performance (administered more shots in 2018 versus 2017), the Ranking clinics outperform their respective Control clinics by 5.4%. Thus, the majority of performance improvement for the Ranking clinics lies with the clinics with negative performance (top left quadrant of Figure 8).

We consider the matched Control clinics as the counter-factual for the performance improvement that the Ranking clinics would have achieved if they had not been exposed to their relative performance. Thus, while the provision of performance feedback seems to have served to elevate the performance of all Ranking clinics, this is particularly true amongst the clinics who might have been the lowest performers. None of the Ranking clinics appear to have entirely given up on the vaccination season. The combination of these findings with the results of Section 7.6 lead us to conclude the presence of Last-Place Aversion may have enticed poor performing clinics to "stick with the program" and attempt to improve throughout the vaccination season. Thus, we find general results consistent with the literature on general performance feedback (Kuhnen and Tymula 2012, Charness et al. 2014) while also illustrating how this improvement can vary in magnitude depending on the *specific* rank or place of subject (Kuziemko et al. 2014). Although in our case the rankings are anonymous, our result is similar to Song et al. (2018) in that the poor performers demonstrate the greatest gains from the introduction of relative performance feedback. We further link this reaction (at least in part) to an aversion for being ranked near last place. This result aligns with the findings of Buell (2019), where the author observes the negative effects of last place (reduced wait satisfaction and increased likelihood of queue abandonment). Furthermore, with regard to improving productivity and performance, the author notes, "…desires to get out of last place, and to avoid falling into last place, are powerful motivators that can help drive human behavior." (Buell 2019) In our context, we find exactly that: an aversion for last place (even when generally defined) drives beneficial behavior and results in performance improvement. Finally, we show that this reaction is a dynamic one and that as different clinics fall in the rankings, they respond by improving in the subsequent week, helping to lead to the overall group improvement.

7.8 Study 2: Limitations

An ongoing question may be whether individual clinics have the capacity to improve (based on local demand conditions). In order to explore this possibility and seek clinic feedback on their overall experience with Compliance Care as a program, we worked with VaxCom to design and distribute a survey to all Compliance Care clinics after the study ended. Based on the feedback from the survey, the findings from Study 2, and the findings of Study 1, we contend that the clinics *do in fact* have spare capacity to administer flu shots. Amongst the treated clinics, all respondents (35 out of 88 clinics responded) confirmed having sufficient capacity to administer additional flu shots. As a result, by focusing attention on flu shots, it was possible to increase the number of shots delivered.

The focusing of attention on flu shots inherently raises the question of where providers are paying *less* attention as a result. We are limited from seeing how other elements of patient health were affected. Nevertheless, we know the CDC and the American Medical Association recommend everyone over 6 months of age get the flu shot (AMA 2018) and, as mentioned above, clinics had enough capacity to administer these shots. Regardless, while we think it is unlikely patient health was negatively affected in other ways, future work should further explore this possibility to validate our supposition.

Another limitation of our study is in the final sample size. Although we used a power analysis to calculate the size of the study, discussions with the company did not permit us to add additional clinics beyond this target. Thus, when attrition occurred, our final sample may have been underpowered and as a result we can only say with 90% certainty that the difference is significant between the Ranking and Control groups when excluding all terminating clinics. Combining the lessons from each sample gives us confidence in the results that we support, but future work should seek to study the topic with a larger sample.

The implementation of the financial incentive highlights another limitation. Specifically, we do not claim that the given structure of 10/15/20% Year-over-Year growth thresholds or the payment amounts of $1/2/3 per shot rebates are optimal. Although it is possible these rebates were too small to serve as significant motivators, we suggest that this is not the case as 75% of the Rebate clinics polled indicated the amount was in fact meaningful. In addition, these dollar amounts correspond to approximately a 5/10/15% bonus on top of the standard shot payment. Nevertheless, future work should identify uniform, optimal thresholds, balancing the desire for growth with an understanding of realistic growth potential.

Another concern is whether our results will generalize from clinics that partner with VaxCom to other healthcare clinics and other industries. In our sample we see significant heterogeneity in clinic location, specialty, capacity, and resulting performance. That the above behaviors manifest in a randomized setting despite this heterogeneity further enhances the robustness of our findings. Future works should continue to explore these topics in healthcare and beyond.

Finally, our identification of a behavioral response makes no effort to explain the mechanism behind the response. We have demonstrated the efficacy of financial incentives and performance feedback, but the reasons underlying the outcomes remain untested. These reasons could include exploring *what* specific behavior the financial incentive did not affect or *why* a clinic exhibits Last Place Aversion. Future work should seek to clarify the processes through which these interventions affect behavior.

8 Discussion & Conclusion

The U.S. Center for Disease Control and Prevention (CDC) is continually looking for a process to reduce the occurrence of influenza outbreaks and lessen the likelihood of an influenza pandemic. While many factors can contribute to a population's ability to prevent these outbreaks, such as improved hand hygiene, vaccines present perhaps the greatest opportunity for large scale improvement. This commitment to improve manifests in the CDC's recommendation that everyone over 6 months of age who is not contraindicated receive the vaccine.[7] Studies find that the results of vaccination are significant. For example, during the 2017-2018 flu season, the CDC estimates the flu vaccine averted almost 7.1 million

[7] See https://www.cdc.gov/flu/consumer/vaccinations.htm for the subset who should not be vaccinated

instances of a flu-like illness and more than 3.7 million medical visits (Rolfes et al. 2019). This is in spite of the fact that during that season the percent of the United States adult population who received the flu vaccine dropped from 43.3% in 2016-17 to 37.1% in 2017-18. The report further demonstrates approximately 1,200 vaccinations must be given in order to prevent one hospitalization. Thus, the call to action is significant and the lack of noticeable gains surrounding flu vaccine coverage in the past decade reemphasizes the opportunity for improvement.

Moving towards compliance with the CDC's recommendation requires participation by two separate parties: health care providers and individual patients. In this paper, we have considered these as the supply and demand sides of the flu vaccine equation. In order for providers to administer the vaccine, patients need to visit their providers office during vaccination season. Most interventions, to date, have focused on a demand side approach to flu vaccination – encouraging patients to visit a clinic to seek out the shot (e.g., Milkman et al. 2011 and Chapman et al. 2016). However, despite individual successes, the flu vaccination rate remains largely unchanged. Perhaps, a better way can be found by investigating a supply side perspective. As Brewer et al. (2017) and Patel et al. (2017) find in their studies of vaccination processes, many patients would be willing to accept the vaccine if their provider simply recommended it. And despite the prevalence of patients seeking a flu vaccine from their local pharmacy (22%), patients still report their primary care provider as the most frequent channel of receiving their flu vaccine (48%, CVS Health 2018).

Our research seeks to address this problem from the supply perspective, focusing on health care providers. We study the introduction of two interventions designed to increase the likelihood of providers recommending the flu vaccine to their patients. The program, Compliance Care, was introduced and piloted by VaxCom, a vaccine management organization. The first Compliance Care intervention (Study 1) targeted Year-over-Year total flu vaccine growth for 101 health care clinics in Florida during the 2017 flu vaccine season. Per-shot financial rebates were provided when clinics achieved various growth thresholds. Additionally, clinics were provided with their ranking versus other clinics who were also in the program. Thus, clinics were motivated by both financial incentives and non-financial feedback to improve their vaccination rate over the course of the 2017 season. The second intervention (Study 2) randomized assignment into two treatment groups and a control group and selected these groups from dispersed geographic locations. Following strict inclusion criteria, 145 clinics were initially recruited to participate in Study 2 and of these, 140 clinics started the 2018 vaccination season as Compliance Care participants. The combined treatments which were present in Study 1 (financial incentives for growth and relative performance feedback) were separated into distinct treatment arms in Study 2.

Our paper makes several notable contributions to the literature. First, we contribute to the growing body of work in operations that finds an important role for discretion by operational actors in driving operational performance (van Donselaar et al. 2010, Campbell and Frei 2011, Kim et al. 2015, Phillips et al. 2015, Ibanez et al. 2017, Ibanez and Toffel 2019, Freeman et al. 2017, Song et al. 2018). We found a supply-side approach was operationally effective: the introduction of Compliance Care led to an increase in vaccines for the clinics in the program. In Study 1, although the total number of vaccinations for the state of Florida was down 9.3% for the clinics outside the program, the most conservative subset of CC clinics was approximately flat in the number of shots given between 2016 and 2017, for a net gain of about 9%. In Study 2, we find a net 9.8% growth in the Ranking clinics compared to the control condition. Our approach of focusing on vaccine *provision*, rather than on vaccine *solicitation*, yielded gains in a domain bereft of any notable change despite years of research and policy attention. We propose this is due to an attention cue induced by the intervention. Future research should seek to clearly identify the mechanisms behind the change in provider behavior. Moreover, additional research is necessary to consider how to better design operating systems to help allocate attention on the right tasks to improve performance.

Second, we highlight the motivational power of social comparison (Festinger 1954, Cole, Mailath, and Postlewaite 1992, Raleigh et al. 1984, Coffey and Maloney 2010), even over and above economic incentives. We find that providing performance feedback dominates the effect of providing financial incentives. Given the literature's proven acceptance of financial incentives (Prendergast 1999, Lazear 2000, Shearer 2004), the mixed results of performance feedback (Dechenaux et al. 2015), and the positive results from Study 1, we expected the Rebate group to outperform both the Control and the Ranking groups. Neither happened in this case. This counter-intuitive result highlights an opportunity to leverage low cost interventions to effectively focus healthcare provider attention on an issue of interest. In addition to structuring rankings in order to improve performance, as discussed below, more research is also needed to explore the construction of financial incentives. Future work should seek to establish robust thresholds which balance incentivizing growth with realistic outcome expectations. The additional possibility of an interaction between performance feedback and financial incentives as in Study 1 presents another intriguing opportunity to further extend our findings.

Third, when considering the clinic rankings and the role of performance feedback, we also find evidence of Last-Place Aversion, similar to Kuziemko et al. (2014), Gill et al. (2019), and Buell (2019). Although such phenomena have been robustly demonstrated in a laboratory setting, to the best of our knowledge, our analysis is the first to examine the presence of these behaviors in productivity in a dynamic, real-world setting. With significant demands on provider time and attention it is not clear whether a ranking with respect to an unknown comparison group will actually drive behavioral change. Not only does the presence of relative performance yield modest growth for all clinics but the effects appear to be the strongest for the worst performing clinics (similar to Song et al. 2018).

In our setting, it appears an aversion for last place induces a clinic to continue administering vaccinations even when the hope for growth appears beyond reach. Consider the comparison of the worst performing Ranking clinics with their respective Control clinics. While the low-performing Control clinics show a significant decline in vaccination rates and appear to have given up on flu vaccinations by the end of the season, the provision of performance feedback to the Ranking clinics results in only modest declines in their vaccinations and a sizable performance gap between these two groups. Future work should further clarify the presence of Last-Place Aversion as well as explore the mechanisms for why clinics exhibit such behavior when there are no negative repercussions for poor performance. Separately, although prior work has robustly established First-Place Loving behavior in the experimental laboratory, we do not find conclusive evidence of its presence in our field experiment. Nevertheless, our study does not permit us to conclude that FPL does not exist – it is possible that it is frame (i.e., rank) dependent or our study may be under-powered to identify the effect. More work is needed to investigate this behavior in the field.

Finally, as a general contribution, we demonstrate the power and importance of running field experiments when addressing complicated operational questions (Ibanez and Staats 2018). Although the use of archival data can yield rich insights, in this instance, the sole application of archival data analysis to Study 1 would have yielded limited (if not biased) results and conclusions. For example, it is clear in Study 1 that Compliance Care resulted in flu shot growth, but the identification of the mechanism would have been unclear: was the growth due to clinic selection, or clinic location, or financial incentives, or performance feedback, or some interaction of a number of these factors? Understanding endogeneity and causal identification in archival studies is important for just this reason. However, such work should not be completely dismissed, but rather serve as motivation for further work to disentangle complicated effects. We do just that with our field experiment. Study 2 tests these possibilities and clearly identifies the strongest driver of clinic growth in this setting. As KC et al. (2019) note "Our field studies many things, but work is central to what we do," and since field experiments present "…the opportunity for researchers to situate their work clearly in the context in which it is actually occurring", researchers who seek to do rigorous and relevant research in operations management should continue to leverage the explanatory power and identification advantages of field experiments moving forward.

8.1 Managerial Implications

Our findings have a number of managerial implications. Although many organizational interventions lead to performance improvement, it is important to note Compliance Care achieved notable growth in administering flu shots, a domain lacking significant improvement in more than a decade in the overall US population. It is also apparent the provision of relative feedback induced competition amongst clinics, even when applied without a financial incentive. In turn, this resulted in an increase in flu shots, particularly among the poorest performing clinics who might have been most likely to give-up on the program. Being mindful of the application of such low-cost incentives could yield performance gains even in settings with financial constraints.

Additionally, when designing the operational structure of an organization to facilitate growth (Hayes and Clark 1988), managers should be mindful of the "reference point structure" as discussed in Roels and Su (2014). Managers have the flexibility to use either (or both) treatment(s) to encourage compliance. In our case, the incentive thresholds and other program participants serve as reference points and provide self-evaluative feedback on current performance (Festinger 1954).

In a similar vein, the operational structure can be leveraged to selectively focus decision makers' attention. In the context of healthcare, the demands on a provider's time and attention are ever increasing, but this finding is not new, and healthcare is not the only context where inattention may lead to sub-optimal decision making. Mintzberg (1973) observed managers constantly working at a frenetic pace decades ago and forecasted these demands would only grow over time. And March and Simon (1992) recognized a firm's decision makers selectively apply attention (or inattention) to organizational issues, depending both on their decision-making capacity and any relevant "time pressure." In this paper, we have shown that in the presence of increasing demands on managers, small attention cues can serve to highlight areas for improvement and encourage these decision makers to selectively focus their attention on an issue of interest. This finding has broad implications which extend beyond healthcare to nearly any organizational context.

Finally, although we have demonstrated a context where an intervention cued the provider's attention to an issue of interest, it is important to note that one cannot possibly cue attention to all desired behaviors. As noted in a physician survey (The Physician's Foundation 2018), most providers are already operating at capacity; we have simply intervened to cue attention to where some capacity remains in the chain of healthcare delivery. And yet, the introduction of multiple attention cues would likely lead to a reduction in the effectiveness of each cue with an end state potentially worse than the initial state. Thus, we caution managers in such contexts to be mindful of the use of incentives and performance feedback and to prioritize interventions which achieve the greatest benefit via the smallest impact, leaving providers free to do that which matters most: provide quality care for their patients.

8.2 Conclusion

All parties stand to benefit by increasing flu vaccination levels. Individuals are less likely to contract the flu, meaning fewer days of missed work and fewer medical visits. Insurance companies are heavily invested in preventative medicine, with significant preferences for patient vaccination over patient hospitalization. Vaccine manufacturers benefit from increased production and any associated economies of scale. And finally, health care providers benefit from the overall improved health of their patient population and a reduction in the demand for their time to treat a preventable illness. Altogether, in accord with Taylor (1911), we conclude appropriate supply-side behavioral intervention programs can improve performance, even when targeting seemingly immutable trends, such as influenza vaccination rates. Our results show promise for noticeable improvements, specifically in public health in the United States and around the world, and more generally for company operations.

9 References

Allcott H (2011) Social norms and energy conservation. *J. Public Econ.* 95:1082–1095.

Allcott H, Rogers T (2014) The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation. *Am. Econ. Rev.* 104(10):3003–3037.

American Medical Association (2018) AMA encourages people 6 months and older to get vaccinated against flu. Retrieved (August 9, 2019), https://www.ama-assn.org/press-center/ama-statements/ama-encourages-people-6-months-and-older-get-vaccinated-against-flu.

Arifoğlu K, Deo S, R Iravani SM, Arifo K (2012) Consumption Externality and Yield Uncertainty in the Influenza Vaccine Supply Chain: Interventions in Demand and Supply Sides. *Manage. Sci.* 58(6):1072–1091.

Ayres I, Raseman S, Shih A (2013) Evidence from Two Large Field Experiments that Peer Comparison Feedback Can Reduce Residential Energy Usage. *J. Law, Econ. Organ.* 29(5):992–1022.

Bandiera O, Barankay I, Rasul I (2013) Team Incentives: Evidence from a Firm Level Experiment. *J. Eur. Econ. Assoc.*

Barankay I (2011) Rankings and Social Tournaments: Evidence from a Field Experiment.

Barankay I (2012) Rank Incentives Evidence from a Randomized Workplace Experiment.

Brewer NT, Chapman GB, Rothman AJ, Leask J, Kempe A (2017) Increasing Vaccination: Putting Psychological Science Into Action. *Psychol. Sci. Public Interes.* 18(3):149–207.

Buell RW (2019) Last Place Aversion in Queues.

Campbell D, Frei F (2011) Market Heterogeneity and Local Capacity Decisions in Services. *Manuf. Serv. Oper. Manag.* 13(1).

CDC (2018) Estimates of Influenza Vaccination Coverage among Adults 2017/18 Flu Season. Retrieved (December 19, 2018), https://www.cdc.gov/flu/fluvaxview/coverage-1718estimates.htm.

CDC (2019) Estimated Influenza Illnesses, Medical visits, Hospitalizations, and Deaths in the United States — 2017–2018 influenza season. Retrieved (July 9, 2019), https://www.cdc.gov/flu/about/burden/estimates.htm.

Chapman GB, Li M, Colby H, Yoon H (2010) Opting In vs Opting Out of Influenza Vaccination. *J. Am. Med. Assoc.* 304(1):43.

Chapman GB, Li M, Leventhal H, Leventhal E (2016) Default clinic appointments promote influenza vaccination uptake without a displacement effect. *Behav. Sci. Policy*.

Charness G, Masclet D, Villeval MC (2014) The Dark Side of Competition for Status. *Manage. Sci.* 60(1):38–55.

Cho SH (2010) The Optimal Composition of Influenza Vaccines Subject to Random Production Yields. *Manuf. Serv. Oper. Manag.* 12(2):256–277.

Coffey B, Maloney MT (2010) The Thrill of Victory: Measuring the Incentive to Win. *J. Labor Econ.* 28(1):87–112.

Cole H, Mailath G, Postlewaite A (1992) Social Norms, Savings Behavior, and Growth. *J. Polit. Econ.*

Cortina JM, Landis RS (2011) The earth is not round (p = .00). *Organ. Res. Methods* 14(2):332–349.

Costa DL, Kahn ME (2013) Energy Conservation "Nudges" And Environmentalist Ideology: Evidence From a Randomized Residential Electricity Field Experiment. *J. Eur. Econ. Assoc.* 11(3):680–702.

CVS Health (2018) CVS Health Survey Finds Consumers More Likely to Get Vaccinated Following Last Year's Record Breaking Flu Season. Retrieved (May 28, 2019), https://cvshealth.com/newsroom/press-releases/cvs-health-survey-finds-consumers-more-likely-get-vaccinated-following-last.

Dai T, Akan M, Tayur S (2017) Imaging Room and Beyond: The Underlying Economics Behind Physicians. *Manuf. Serv. Oper. Manag.* 19(1):99–113.

Dechenaux E, Kovenock D, Sheremeta RM (2015) A survey of experimental research on contests, all-pay auctions and tournaments. *Exp. Econ.* 18:609–669.

Deci EL (1971) Effects of externally mediated rewards on intrinsic motivation. *J. Pers. Soc. Psychol.* 18(1):105–115.

Deo S, Corbett CJ (2009) Cournot Competition Under Yield Uncertainty: The Case of the U.S. Influenza Vaccine Market. *Manuf. Serv. Oper. Manag.* 11(4):563–576.

Eriksson T, Poulsen A, Villeval MC (2009) Feedback and incentives: Experimental evidence. *Labour Econ.* 16:679–688.

FDA (2019) CBER-Regulated Products: Shortages and Discontinuations - CBER-Regulated Products: Resolved Shortages. Retrieved (August 20, 2018), https://www.fda.gov/BiologicsBloodVaccines/SafetyAvailability/Shortages/ucm351943.htm.

Fehr E, Gächter S (2000) Fairness and Retaliation: The Economics of Reciprocity. *J. Econ. Perspect.*

Festinger L (1954) A Theory of Social Comparison Processes. *Hum. Resour.*

Freeman M, Savva N, Scholtes S (2017) Gatekeepers at Work: An Empirical Analysis of a Maternity Unit. *Manage. Sci.* 63(10):3147–3167.

Frey BS (1993) Shirking or work morale? *Eur. Econ. Rev.* 37:1523–1532.

Garten R, Blanton L, Abd Elal Isa A, Alabi N, Barnes J, Biggerstaff M, Brammer L, et al. (2018) *The United States 2017-18 influenza season*

Gill D, Kissová Z, Lee J, Prowse V (2019) First-place loving and last-place loathing: How rank in the distribution of performance affects effort provision. *Manage. Sci.*

Gneezy U, Meier S, Rey-Biel P (2011) When and Why Incentives (Don't) Work to Modify Behavior. *J. Econ. Perspect.* 25(4):191–210.

Hayes RH, Wheelwright SC, Clark KB (1988) *Dynamic Manufacturing: Creating the Learning Organization* (New York, Free Press.).

Huckman RS, Zinner DE (2008) Does Focus Improve Operational Performance? Lessons from the Management of Clinical Trials. *Strateg. Manag. J.* 29:173–193.

Ibanez MR, Huckman RS, Clark JR, Staats BR (2017) Discretionary Task Ordering: Queue Management in Radiological Services. *Manage. Sci.*

Ibanez MR, Staats BR (2019) CH 4 - Behavioral Empirics and Field Experiments. *Handb. Behav. Oper. Manag.*

Ibanez MR, Toffel MW How Scheduling Can Bias Quality Assessment: Evidence from Food Safety Inspections. *Manage. Sci.*

KC DS, Staats BR, Kouchaki M, Gino F (2019) Task Selection and Workload: A Focus on Completing Easy Tasks Hurts Performance How. *Manage. Sci.*:1–39.

Kim SH, Chan CW, Olivares M, Escobar G (2015) ICU Admission Control: An Empirical Study of Capacity Allocation and Its Implication for Patient Outcomes. *Manage. Sci.* 61(1):19–38.

Kuhnen CM, Tymula A (2012) Feedback, Self-Esteem, and Performance in Organizations. *Manage. Sci.* 58(1):94–113.

Kuziemko I, Buell RW, Reich T, Norton MI (2014) "Last-Place Aversion": Evidence and Redistributive Implications. *Q. J. Econ.*

Lazear EP (2000) Performance Pay and Productivity. *Am. Econ. Rev.* 90(5):1346–1361.

Legge S, Schmid L (2015) Media attention and betting markets. *Eur. Econ. Rev.*

March J, Simon HA (1992) *Organizations* (Blackwell Publishers).

Meeker D, Linder JA, Fox CR, Friedberg MW, Persell SD, Goldstein NJ, Knight TK, Hay JW, Doctor JN (2016) Effect of behavioral interventions on inappropriate antibiotic prescribing among primary care practices a randomized clinical trial. *J. Am. Med. Assoc.* 315(6):562–570.

Milkman KL, Beshears J, Choi JJ, Laibson D, Madrian BC (2011) Using implementation intentions prompts to enhance influenza vaccination rates. *Proc. Natl. Acad. Sci. United States Am.*

Mintzberg H (1973) *The Nature of Managerial Work* (New York, Harper and Row).

Molinari NAM, Ortega-Sanchez IR, Messonnier ML, Thompson WW, Wortley PM, Weintraub E, Bridges CB (2007) The annual impact of seasonal influenza in the US: Measuring disease burden and costs. *Vaccine* 25:5086–5096.

Pasternack BA (1985) Optimal Pricing and Return Policies for Perishable Commodities. *Mark. Sci.* 4(2):166–176.

Patel MS, Volpp KG, Small DS, Wynne C, Zhu J, Yang L, Honeywell S, Day SC (2017) Using Active Choice Within the Electronic Health Record to Increase Influenza Vaccination Rates. *J. Gen. Intern. Med.* 32(7):790–795.

Phillips R, Serdar Şimşek A, van Ryzin G, Serdar A (2015) The Effectiveness of Field Price Discretion: Empirical Evidence from Auto Lending. *Manage. Sci.* 61(8):1741–1759.

Prendergast C (1999) The Provision of Incentives in Firms. *J. Econ. Lit.* 37(1):7–63.

Raleigh MJ, Mcguire MT, Brammer GL, Yuwiler A (1984) Social and Environmental Influences on Blood Serotonin Concentrations in Monkeys. *Arch. Gen. Psychiatry* 41(4):405–410.

Roels G, Su X (2014) Optimal Design of Social Comparison Effects: Setting Reference Groups and Reference Points. *Manage. Sci.* 60(3):606–627.

Rolfes MA, Flannery B, Chung JR, O'halloran A, Garg S, Belongia EA, Gaglani M, et al. (2019) Effects of Influenza Vaccination in the United States During the 2017-2018 Influenza Season. *Clin. Infect. Dis.*

Shearer B (2004) Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment. *Rev. Econ. Stud.* 71(2):513–534.

Simmons J, Nelson L, Simonsohn U (2012) A 21 Word Solution. *Dialogue Can. Philos. Assoc.* http://ssrn.com/abstract=2160588.

Simmons JP, Nelson LD, Simonsohn U (2011) False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychol. Sci.* 22(11):1359–1366.

Simon HA (1947) *Administrative Behavior: A Study of Decision-making Processes in Administrative Organizations* (Macmillan, Chicago, IL).

Song H, Tucker AL (2016) *Performance Improvement in Health Care Organizations*

Song H, Tucker AL, Murrell KL, Vinson DR (2018) Closing the Productivity Gap: Improving Worker Productivity Through Public Relative Performance Feedback and Validation of Best Practices. *Manage. Sci.* 64(6):2628–2649.

Staats BR, Dai H, Hofmann D, Milkman KL (2017) Motivating Process Compliance Through Individual Electronic Monitoring: An Empirical Examination of Hand Hygiene in Healthcare. *Manage. Sci.*

Taylor FW (1911) *The Principles of Scientific Management*

The Physician's Foundation (2018) 2018 Survey of America's Physicians. Retrieved (May 6, 2019), https://physiciansfoundation.org/research-insights/the-physicians-foundation-2018-physician-survey/.

Tsai TC, Jha AK, Gawande AA, Huckman RS, Bloom N, Sadun R (2015) Hospital Board And Management Practices Are Strongly Related To Hospital Performance On Clinical Quality Metrics. *Health Aff.*

Tucker A (2007) An Empirical Study of System Improvement by Frontline Employees in Hospital Units. *Manuf. Serv. Oper. Manag.*

van Donselaar KH, Gaur V, van Woensel T, Broekmeulen RACM, Fransoo JC (2010) Ordering Behavior in Retail Stores and Implications for Automated Replenishment. *Manage. Sci.* 56(5):766–784.

Xu X, Blanton L, Elal AAI, Alabi N, Barnes J, Biggerstaff M, Brammer L, et al. (2019) *Update: Influenza Activity in the United States During the 2018–19 Season and Composition of the 2019–20 Influenza Vaccine*

10 Tables

Table 1: Number of clinics who achieved each incentive tier in Study 1 (2017) and Study 2 (2018)

	Study 1: 2017	Study 2: 2018
No Tier Achieved	63	31
Achieved Tier 1	3	5
Achieved Tier 2	3	3
Achieved Tier 3	32	9
Total #	101	48

Note: All clinics in Study 1 were provided with incentive tiers and the opportunity to earn financial rebates for exceeding these growth tiers (N = 101). In Study 2, only the Rebate clinics were provided with incentive tiers and the opportunity to earn financial rebates for exceeding these growth tiers. For Study 2, the results are presented for the clinics from the "Partial Exclusion" subset (N = 48).

Table 2: Study 1 - Flu shot counts and percent growth in Florida between 2016 and 2017

	# Clinics	2016	2017	% Growth	Net Effect
Non-CC Clinics					
- YOY Partners	68	10,947	9,933	-9.26%	-
CC Clinics					
- All CC Clinics	101	22,295	27,707	24.27%	33.54%
- YOY Partners	23	7,133	7,110	-0.32%	8.94%

Note: This table compares clinics in Compliance Care 2017 to other Florida clinics outside of Compliance Care 2017. The Year-over-Year (YOY) Partners fully participated in the 2016 and 2017 flu shot seasons. The remaining clinics in Florida who were outside of CC were ignored for this analysis due to their unknown performance in 2016. The "Net Effect" is the difference between a group's percent growth and the Non-CC clinics' percent growth.

Table 3: Study 2 - Flu shot counts and percent growth between 2017 and 2018 for the "Full" clinic set

	# Clinics	2017	2018	% Growth	Net Effect
Control	49	18,546	18,301	-1.32%	-
Ranking	47	17,529	19,495	11.22%	12.54%
Rebate	49	16,955	17,315	2.12%	3.44%

Note: These clinics were randomly assigned to either the Control, Ranking, or Rebate group and agreed to participate in Compliance Care 2018 before the start of the flu vaccine season. As such, this grouping includes some clinics who later terminated partnership with VaxCom. We refer to this as the "Full" clinic set (N = 145). The "Net Effect" is the difference in percent growth between the Ranking/Rebate group and the Control.

Table 4: Study 2 - Flu shot counts and percent growth between 2017 and 2018 for the "Partial Exclusion" subset of clinics

	# Clinics	2017	2018	% Growth	Net Effect
Control	46	17,894	18,301	2.27%	-
Ranking	46	17,389	19,495	12.11%	9.84%
Rebate	48	16,776	17,315	3.21%	0.94%

Note: These clinics were randomly assigned to either the Control, Ranking, or Rebate group, agreed to participate in CC 2018, and started the vaccination season as VaxCom partners in Compliance Care 2018. As such, this grouping includes some clinics who later terminated partnership with VaxCom during the vaccination season. We refer to this as the "Partial Exclusion" subset of clinics (N = 140). The "Net Effect" is the difference in percent growth between the Ranking/Rebate group and the Control.

Table 5: Regression Model – Effect of Compliance Care 2018 treatments (Rankings/Rebates) on *Shots per Patient Population* (DV1) for the "Partial Exclusion" clinic subset

	Shots per Patient Population (1)	**Shots per Patient Population (2)**
RE/FE Model	Fixed Effects	Random Effects
Treatment: Ranking		
- Coefficient	0.0458**	0.0458+
- Standard Errors	(0.0231)	(0.0234)
- p-value	0.0495	0.0503
Treatment: Rebate		
- Coefficient	0.0109	0.0109
- Standard Errors	(0.0237)	(0.0240)
- p-value	0.647	0.650
Observations	5,880	5,880
# clinics	140	140
Years	2017, 2018	2017, 2018
State FE	-	Y
Clinic FE	-	Y

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

Note: This table details the specific effects of the treatment in Study 2 on a clinic's Shots per Patient Population (SPP) for a Fixed Effects and a Random Effects model on the "Partial Exclusion" clinic subset (N = 140). This can be interpreted as the Ranking treatment led to approximately 0.046 *more* shots per clinic patient, or that the Ranking clinics vaccinated approximately 4.6% more of their patient population than the control clinics. The Rebate treatment shows no statistically significant effect. Note the p-value for the Ranking group in column 2 just exceeds the 0.05 threshold. The errors for these models were clustered at the clinic level. The FE model controls for clinic fixed effects by design and drops state fixed effects due to their time-invariant nature.

Table 6: Regression Model – Robustness check of the effect of Compliance Care 2018 treatments (Rankings/Rebates) on *Shots per Patient Population* (DV1) for the "Full" and "Exclusive" clinic subsets

	Shots per Patient Population (1)	**Shots per Patient Population (2)**	**Shots per Patient Population (3)**	**Shots per Patient Population (4)**
RE/FE Model	Fixed Effects	Random Effects	Fixed Effects	Random Effects
Treatment: Ranking				
- Coefficient	0.0655**	0.0655**	0.0359	0.0359
- Standard Errors	(0.0258)	(0.0261)	(0.0219)	(0.0222)
- p-value	0.0123	0.0123	0.104	0.105
Treatment: Rebate				
- Coefficient	0.0272	0.0272	-0.000985	-0.000985
- Standard Errors	(0.0270)	(0.0273)	(0.0224)	(0.0226)
- p-value	0.316	0.320	0.965	0.965
Observations	6,090	6,090	5,670	5,670
# clinics	145	145	135	135
Clinic Set	Full	Full	Exclusive	Exclusive
State FE	-	Y	-	Y
Clinic FE	-	Y	-	Y

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

Note: This table details the specific effects of the treatment in Study 2 on a clinic's Shots per Patient Population (SPP) for a Fixed Effects and a Random Effects models on the "Full" (N = 145) and "Exclusive" (N = 135) clinic subset. This can be interpreted as the Ranking treatment led to approximately 0.065 *more* shots per clinic patient in the Full clinic set, or that the Ranking clinics vaccinated approximately 6.5% more of their patient population than the control clinics in the full clinic set. Note the p-values in columns 3 and 4 for the Ranking group just exceed the 0.10 threshold. The Rebate treatment shows no statistically significant effect in any of the clinic sets in any of the models. The errors for these models were clustered at the clinic level. The FE model controls for clinic fixed effects by design and drops state fixed effects due to their time-invariant nature.

Table 7: Regression Model – Effect of Compliance Care 2018 treatments (Rankings/Rebates) on clinic *Percent Growth* (DV2) for the "Partial Exclusion" clinic subset

	Percent Growth (1)
RE/FE Model	Random Effects
Treatment: Ranking	
- Coefficient	0.0793**
- Standard Errors	(0.0395)
- p-value	0.0446
Treatment: Rebate	
- Coefficient	0.00398
- Standard Errors	(0.0396)
- p-value	0.920
Observations	2,800
# clinics	140
Clinic Set	Partial Exclusion
Patient Count Controls	Y
State FE	Y

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

Note: This table details the specific effects of the treatment in Study 2 on a clinic's percent growth in 2018 for a Random Effects model on the "Partial Exclusion" clinic subset (N = 140). Because this model only examines percent growth in 2018, we cannot evaluate this model with a Fixed Effects model due to the time-invariant nature of the treatment (clinic status, ie, Ranking/Rebate/Control, does not change in 2018). The results show that, on average, the Ranking clinics grew by approximately 7.9% more in their shots than the Control clinics while controlling for clinic patient count and state fixed effects. The Rebate clinics shows no statistically significant difference in growth from the Control clinics. The errors for this model were clustered at the clinic level.

Table 8: Regression Model Sensitivity Analysis – Iterating the definition for "Last Place," and the effect of moving into "Last Place" on a *weekly change in shots per patient* (SPPΔ, DV3) and a *weekly change in shots given as a percent of clinic total* (PCTΔ, DV4)

	SPPΔ (1)	SPPΔ (2)	SPPΔ (3)	SPPΔ (4)	SPPΔ (5)	SPPΔ (6)	SPPΔ (7)	SPPΔ (8)	SPPΔ (9)	SPPΔ (10)	SPPΔ (11)
Variable Label	DV3	DV3	DV3	DV3	DV3	DV3	DV3	DV3	DV3	DV3	DV3
Low Flag = 1 if Rank:	>= 36	>= 37	>= 38	>= 39	>= 40	>= 41	>= 42	>= 43	>= 44	>= 45	>= 46
Coefficient: Low Rank	0.00991	0.0137**	0.0168***	0.0173***	0.0231***	0.0242***	0.0275***	0.0309***	0.0368***	0.0411***	0.0331***
- Standard Errors	(0.00629)	(0.00612)	(0.00577)	(0.00600)	(0.00635)	(0.00671)	(0.00773)	(0.00944)	(0.00872)	(0.0116)	(0.00604)
- p-value	0.122	0.0299	0.00560	0.00618	0.000709	0.000779	0.000883	0.00204	0.000119	0.000960	1.78e-06
Week FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	792	792	792	792	792	792	792	792	792	792	792
# clinics	46	46	46	46	46	46	46	46	46	46	46

	PCTΔ (1)	PCTΔ (2)	PCTΔ (3)	PCTΔ (4)	PCTΔ (5)	PCTΔ (6)	PCTΔ (7)	PCTΔ (8)	PCTΔ (9)	PCTΔ (10)	PCTΔ (11)
Variable Label	DV4	DV4	DV4	DV4	DV4	DV4	DV4	DV4	DV4	DV4	DV4
Low Flag = 1 if Rank:	>= 36	>= 37	>= 38	>= 39	>= 40	>= 41	>= 42	>= 43	>= 44	>= 45	>= 46
Coefficient: Low Rank	0.0161	0.0210+	0.0234**	0.0235**	0.0321***	0.0342***	0.0396***	0.0423**	0.0462***	0.0533**	0.0420***
- Standard Errors	(0.0121)	(0.0123)	(0.0107)	(0.0104)	(0.0105)	(0.0114)	(0.0132)	(0.0162)	(0.0154)	(0.0214)	(0.00976)
- p-value	0.192	0.0939	0.0334	0.0284	0.00363	0.00446	0.00425	0.0122	0.00425	0.0166	8.95e-05
Week FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	792	792	792	792	792	792	792	792	792	792	792
# clinics	46	46	46	46	46	46	46	46	46	46	46

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

Note: This table details the specific effects Ranking clinics moving into various definitions of last place. As an example, Column 7 defines last place as any of the bottom 5 ranks (42, 43, 44, 45, 46 since all are greater than or equal to 42) and moving into these ranks leads to a +2.75% change in shots per patient population (DV3) the following week. Column 1 illustrates where the effect disappears, statistically speaking. There are 46 clinics in the Partial Exclusion subset, and so the bottom rank is Rank 46. All models were evaluated as Fixed Effects models, and the errors for all models were clustered at the clinic level.

11 Figures

Figure 1: From Garten et al. (2018) - Cumulative rates of hospitalization for laboratory-confirmed influenza by week each season

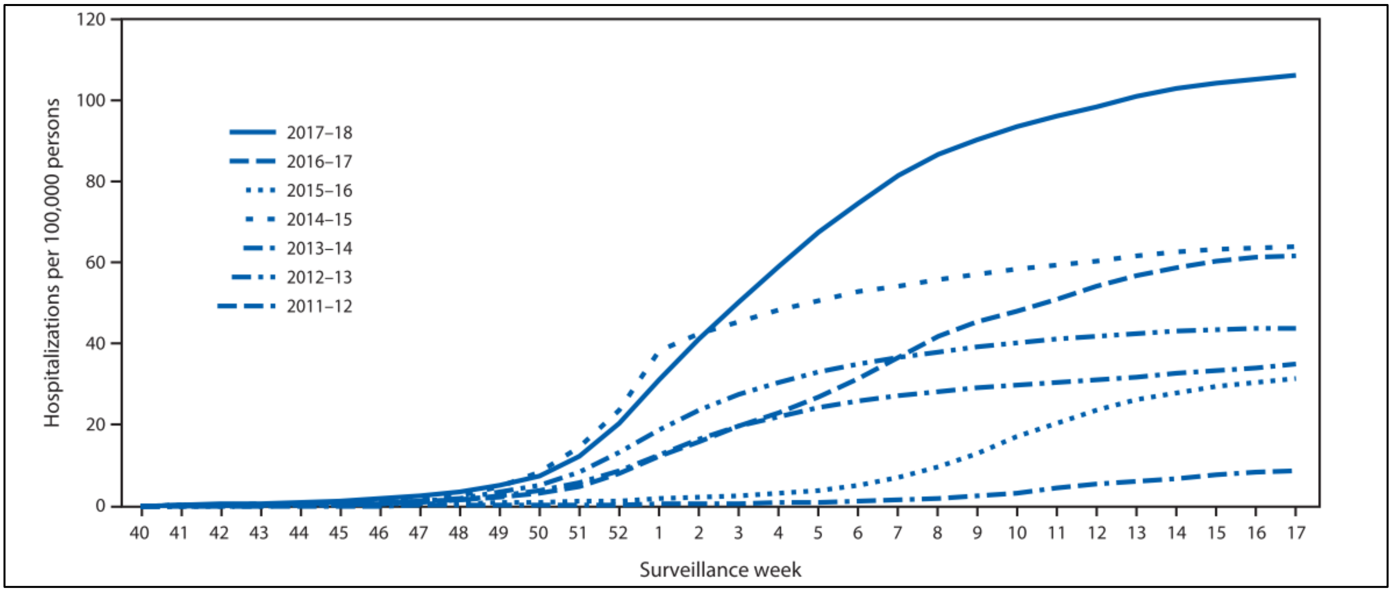
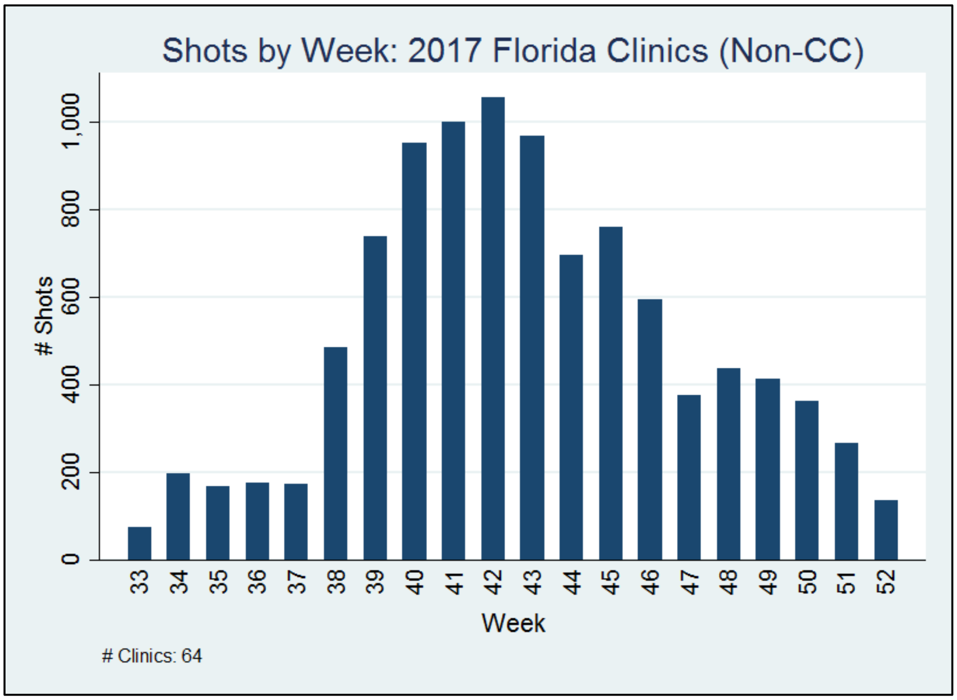
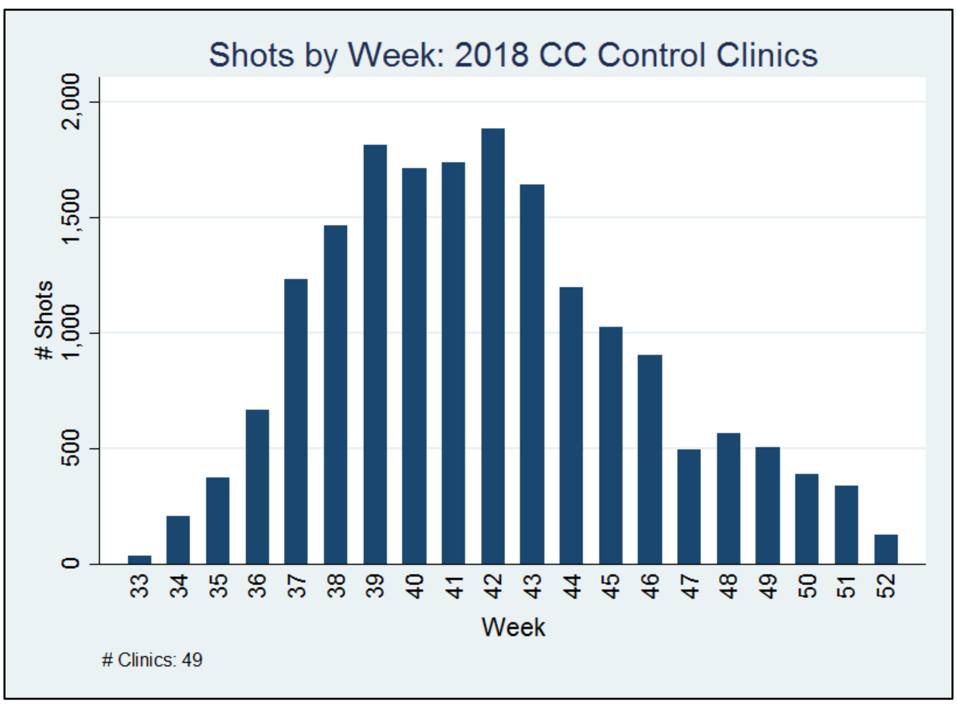


Figure 2: Progression of total shots for Florida Non-CC Clinics in 2017



Note: Week 33 starts with 14 August. Week 40 begins on 2 October.

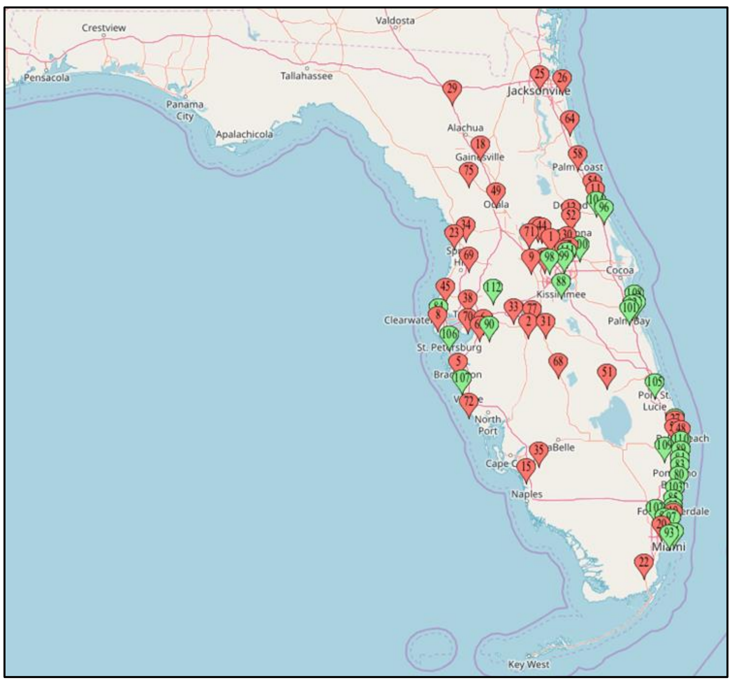
Figure 3: Progression of total shots for Control Clinics in 2018



Note: Week 33 starts with 13 August. Week 40 begins on 1 October.

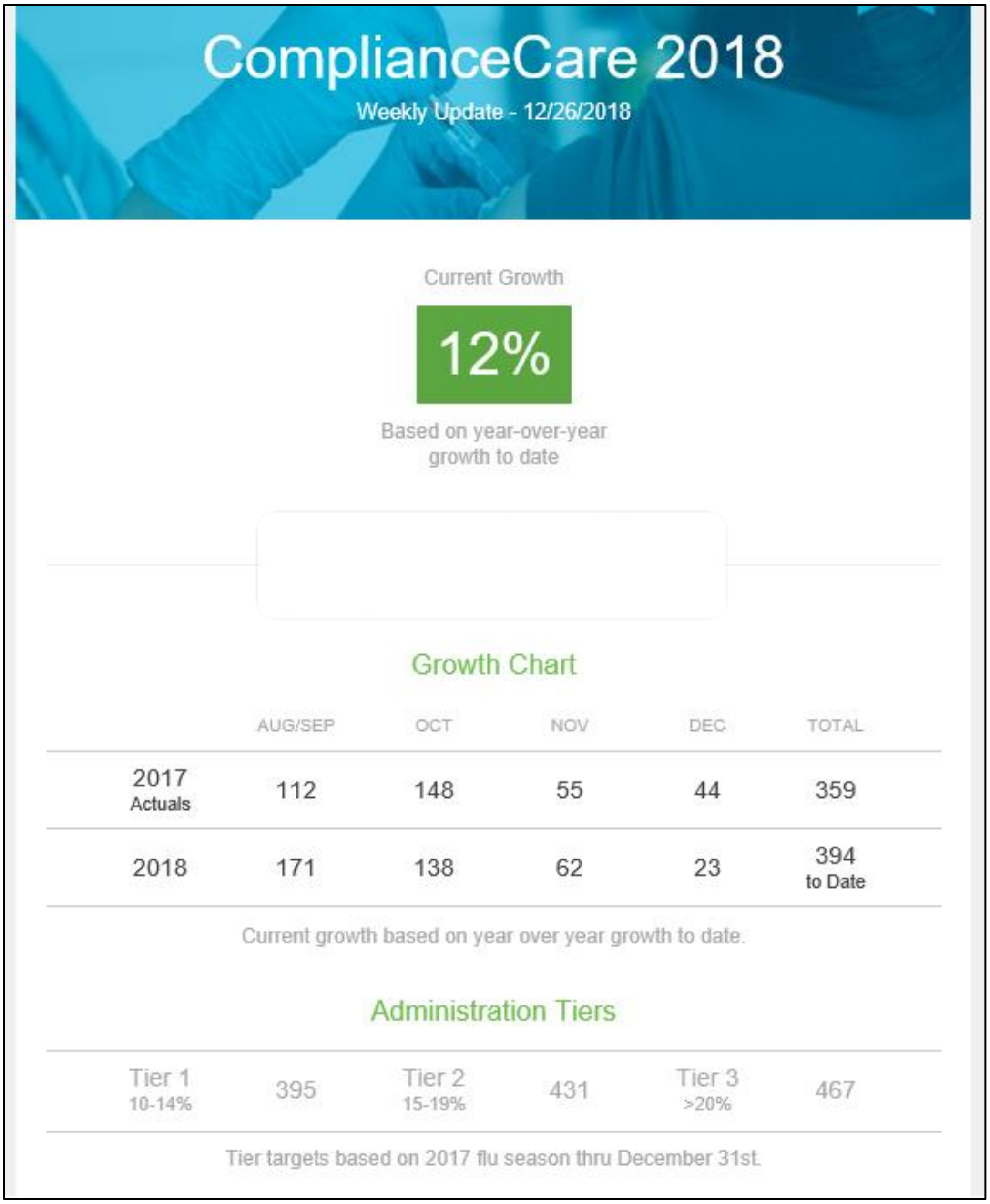
Figure 4: VaxHub Terminal, including UPC scanner and example vaccination vial

Figure 5: Map of Florida clinics



Note: Green markers indicate Compliance Care 2017 clinics. Red markers indicate Non-CC clinics. Clinics with similar locations may appear co-located.

Figure 6: Weekly Status Template – Rebate Group



ComplianceCare 2018

Weekly Update - 12/26/2018

Current Growth

12%

Based on year-over-year growth to date

Growth Chart

	AUG/SEP	OCT	NOV	DEC	TOTAL
2017 Actuals	112	148	55	44	359
2018	171	138	62	23	394 to Date

Current growth based on year over year growth to date.

Administration Tiers

Tier 1 10-14%	395	Tier 2 15-19%	431	Tier 3 >20%	467

Tier targets based on 2017 flu season thru December 31st.

Note: The Growth Chart illustrates the shots administered in the vaccination seasons in 2017 and 2018. The Administration Tiers are the rebate incentive thresholds based on Year-over-Year growth from 2017 to 2018.

Figure 7: Weekly Status Template - Ranking Group

Current Ranking

37
Out of 47 Practices

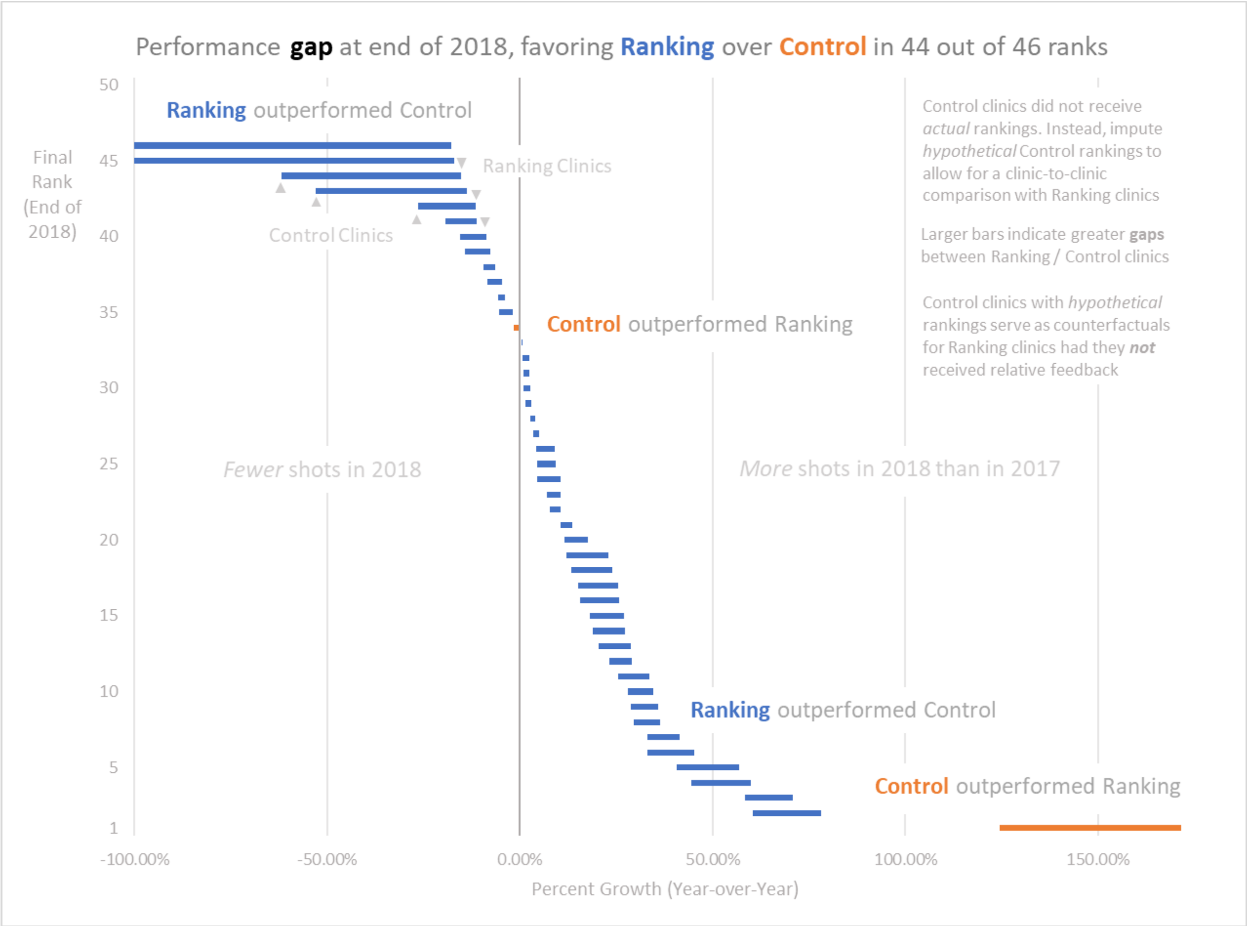
Based on performance
against baseline and peers

Growth Chart

	AUG/SEP	OCT	NOV	DEC	TOTAL
2017 Actuals	1	143	38	12	194
2018	9	116	44	14	183 to Date

Current growth based on year over year growth to date.

Figure 8: Performance gap between Ranking and Control clinics by rank at end of 2018 vaccination season



Note: Higher ranks (ie, 42-46) are worse ranks which correspond to clinics with worse performance, defined along x-axis as Percent Growth, year-over-year. Lower ranks (ie, 1-5) are better ranks which correspond to clinics with better performance, similarly defined along x-axis as Percent Growth year-over-year. As a specific example, consider the comparison of the worst performing Ranking clinics with their respective Control clinics (top left of graph). While the low-performing Control clinics show a significant decline in vaccination rates (left edge of blue bar), the provision of performance feedback to the Ranking clinics results in only modest declines (right edge of blue bar) in their vaccinations and a sizable performance gap (total size of blue bar) between these two groups. In the top left quadrant of the figure, the average gap between Ranking and Control clinics who administered fewer shots in 2018 than in 2017 is 22.9%.

12 Appendix: Detailed Data & Summary Statistics

Table 9: VaxCom Data Fields

Clinic Information	**Shot Information**
Clinic zip code, state	Date Shot Administered, or Date of Service (DOS)
Clinic specialty (e.g. Flu-Only, Pediatric, etc.)	Patient Year of Birth
Date clinic administered first VaxCom vaccination	Patient Gender
Date clinic terminated VaxCom partnership*	Patient Insurance-type
Incentive targets for clinics in CC*	Specific Product Administered

**if applicable*

Table 10: Clinic shot total, number of clinic patients, and average date first flu shot given during season for "Partial Exclusion" clinic subset

		2017	**2018**
Control	Avg. Clinic Shot Total	389	398
	Avg. Clinic Patient Count	739	898
	First Flu-Shot Date	4-Sep-17	2-Sep-18
Ranking	Avg. Clinic Shot Total	378	424
	Avg. Clinic Patient Count	697	803
	First Flu-Shot Date	2-Sep-17	31-Aug-18
Rebate	Avg. Clinic Shot Total	350	361
	Avg. Clinic Patient Count	607	844
	First Flu-Shot Date	4-Sep-17	31-Aug-18

Note: These clinics were randomly assigned to either the Control, Ranking, or Rebate group, agreed to participate in CC 2018, and started the vaccination season as VaxCom partners in Compliance Care 2018. As such, this grouping includes some clinics who later terminated partnership with VaxCom during the vaccination season. We refer to this as the "Partial Exclusion" subset of clinics (N = 140).

Table 11: Clinic average flu shots administered by State & Year & CC Participation

		2017	2018
All States	Non-CC	215	290
	CC 2017	274	366*
	CC 2018	366	399
All States, CC 2018	Control	378	416
	Ranking	373	424
	Rebate	346	361
AL	Non-CC	0	326
CO	Non-CC	217	241
	CC	367	415
FL	Non-CC	162	199
	CC 2017	274	366*
	CC 2018	357	391
GA	Non-CC	216	254
	CC	344	401
IL	Non-CC	213	440
IN	Non-CC	242	253
	CC	435	485
KS	Non-CC	332	424
KY	Non-CC	213	284
	CC	364	372
MO	Non-CC	209	258
	CC	244	259
NY	Non-CC	361	248
OH	Non-CC	262	297
	CC	446	449
OK	Non-CC	0	246
PA	Non-CC	249	461
	CC	368	401
SC	Non-CC	190	315
	CC	395	407
TX	Non-CC	212	569
WV	Non-CC	280	302

Note: Averages include all clinics who administered a non-zero number of shots in the given year. This subset of clinics is referred to as the “Exclusive” clinic subset and includes no clinics who terminated partnership with VaxCom (N = 135).
*VaxCom continued to monitor CC 2017 clinics during the 2018 season.

13 Appendix: Study 1 – Further Robustness

13.1 Sensitivity Analysis of First Place and Last Place Definitions

To determine the robustness of these results to our definitions of first and last place, we can test the sensitivity of these results to other thresholds surrounding the top or bottom 10 ranks. When testing for FPL behavior in all 101 clinics, we continue to find no statistically significant effect when restricting to the top 5 or when relaxing to the top 15 (top 5% and 15% of rankings). When testing for LPA behavior in all 101 clinics, we continue to find a statistically significant effect when restricting to the bottom 9 ranks ($p = 0.0643$) after which the effect continues to be of similar size but not statistically significant ($p = 0.146$, 0.104, and 0.117 when threshold set at bottom 8, 7 and 6, respectively). We also continue to find a statistically significant effect of similar size when relaxing to the top 15 ranks ($p = 0.001$).

14 Appendix: Clinic Distribution (Study 2)

Table 12: Initial Distribution of Clinics after CC 2018 Enrollment

		CO	FL	GA	IN	KY	MO	OH	PA	SC	*sum*
Treatment Arm	Control	6	10	12	3	5	3	6	3	1	49
	Ranking	6	11	12	2	4	2	6	3	1	47
	Rebate	4	11	12	4	5	4	6	3	0	49
							Total # of Clinics:				145

Table 13: Clinic Distribution at Start of Vaccination Season

		CO	FL	GA	IN	KY	MO	OH	PA	SC	*sum*
Treatment Arm	Control	5	10	12	2	5	3	6	2	1	46
	Ranking	6	11	11	2	4	2	6	3	1	46
	Rebate	4	10	12	4	5	4	6	3	0	48
							Total # of Clinics:				140

Table 14: Final Clinic Distribution at Season End

		CO	FL	GA	IN	KY	MO	OH	PA	SC	*sum*
Treatment Arm	Control	4	10	12	2	4	3	6	1	1	43
	Ranking	6	10	11	2	4	2	6	3	1	45
	Rebate	4	10	12	4	4	4	6	3	0	47
							Total # of Clinics:				135

15 Appendix: Further Robustness Tests

15.1 Hausman Test of Random Effects Models

We initially performed the reduced-form analysis controlling for fixed-effects. A common question of interest is whether a random-effect approach might produce more efficient results with less error. One could make the argument each clinic's individual characteristics are randomly drawn from some population. A Hausman test is commonly performed to evaluate the validity of this argument. In a fixed-effect model, the estimator is always consistent, but may not be efficient. A random-effect estimator is efficient, but if the underlying assumption (the clinic specific characteristics are random) is violated, then the estimator will be inconsistent. The Hausman test checks to see if in the probability limit these two estimators converge, with the null hypothesis claiming the two estimators converge; accepting the null would imply the random-effect estimator is consistent and would be preferred over a fixed-effect estimator due to lower variance.

One limitation of the Hausman tests is the clustering of standard errors. In the above analysis, errors were cluster at the clinic level after pooling all observations. A Hausman test requires heteroskedasticity and cannot include time fixed effects. Since the Fixed-Effects model is sufficient when considering the results of Section 7.2 and a Random-Effects model is the only feasible model in Section 7.4, we did not apply any further Hausman tests to the above models.

16 Appendix: Data and Code Disclosure

The data contained in the above paper is subject to a Non-Disclosure Agreement between the aforementioned researchers and VaxCom. Such an agreement prevents the sharing of data due to the sensitive nature of both the business model and the patient information. Here we seek to provide a construct of how the data is available in order to aid replication if one had access to a similar dataset.

The shot records data as captured by VaxCom's technology hub was provided to us with the following information for each shot. For the purposes of this paper, we only focused on flu related products and ignored the remaining vaccine products. We also ignored the payer category, provider identifier, and patient YOB.

- Date of Service
- Patient Year of Birth
- Patient Gender
- Clinic Identifier
  - Unique Count: 2,986
- Provider Identifier
  - Unique Count: 4,230
- Product Administered
  - Unique Count (Flu): 53
- Payer Category
  - Employer Pay, Insurance Pay, Medicare, Medicare Advantage, Multiple, No Pay, Partner Billed, Self-Pay

Separately, we were also provided with specific clinic information (see following). Using the Clinic identifier, we combined these two datasets and then the shot information was aggregated by day or by week (depending on this analysis) for each clinic. The resulting panel dataset included a cumulative count of flu shots given by day (week) for each clinic. As discussed in the paper, we focused only flu shots administered during the flu vaccine season which is between 1 August and 31 December.

- Clinic Identifier
- Partner Identifier
- Clinic State (Variable: $ClinicState_i$)
  - Unique Count: 16
- Zip Code
  - Unique Count: 1,392
- The date that clinic first administered a vaccine through the VaxCom system
- The clinic termination date (if applicable)
- Clinic specialty:
  - Emergency Services, Family Medicine, Health Department, Internal Medicine, OBGYN, Onsite, Pediatric, Unknown
- The clinic baseline for flu shots in 2016 (CC 2017 clinics)

- The clinic baseline for flu shots in 2017
- The clinic baseline for flu shots in 2018
- A binary flag for whether the clinic participated in CC 2017 (Variable: $CCclinic_i$)
- A binary flag for whether the clinic participated in CC 2018
- For CC 2018 clinics, their assigned grouping (Variable: $TreatmentGroup_i$)
    - Control, Ranking, Rebate
- The incentive tiers for a clinic in CC 2017
- The incentive tiers for a clinic in CC 2018 (Rebate group)
- The patient count for a CC 2018 clinic during the flu shot season in 2017 (Variable: $PatientCount_{)}$
- The patient count for a CC 2018 clinic during the flu shot season in 2018 (Variable: $PatientCount_{)}$

Using this dataset, we were able to calculate the following variables of interest:

- Year/Month/Week were derived from the Date of Service
- For clinics in CC 2017, each clinic's cumulative percent progress to Tier 2 was calculated
    - This was the information that was provided to the clinic as well as the information that was used to calculate clinic rankings in 2017
- Calculated CC 2017 clinic rankings by day (week)
    - Range: 1 to 101
- For clinics in CC 2018, each clinic's year-over-year percent growth was calculated exactly as VaxCom calculated it and displayed it to the clinics in the weekly email:
    - This measure was defined as the number of shots given each week divided by the number of shots given that same week the year before minus 1 to give percent growth
    - Prior to Week 49, 2018, this percent growth was rounded to the nearest 10%
    - Starting in Week 49, 2018, this percent growth was no longer rounded
    - If the clinic started vaccinating earlier in 2018 than they had in 2017, the percent growth was simply calculated as the total number of shots given to date
- Given this year-over-year percent growth, the clinics in the Ranking group were assigned ranks identically to the way VaxCom assigned ranks:
    - Range: 1 to 46
    - Ties were assigned the same rank, with an appropriate number of gaps following the number clinics who were tied (similar to track and field rankings)
    - If the clinic started vaccinating later in 2018 than they had in 2017, no rank was assigned until the clinic started vaccinating
    - This Ranking was verified by comparing with VaxCom's original CC data
    - We used the identical process to calculate the hypothetical ranks for the control group
- Terminating clinics were assigned a binary flag (set equal to 1) if:
    - They terminated before the vaccine season began and so saw zero patients that season (ie, patient count mentioned above was equal to zero in 2018)
    - They terminated at any time before the end of the vaccine season
    - Separate analyses for each of these subsets of clinics were run as discussed in the paper (see discussion on Full, Partial Exclusion, and Exclusive clinic sets in Section 6.6)

- Each treatment group was treated as a factor variable
    - Control: 1
    - Ranking: 2
    - Rebate: 3
- Each state with Compliance Care clinics was treated as a factor variable
    - States: CO, FL, GA, IN, KY, MO, OH, PA, SC
    - Assigned integer values 1 through 9
- The following dependent variables were constructed as follows:
    - Shots per clinic patient was calculated for each clinic as cumulative number of flu shots given divided by the total number of patients a clinic saw that vaccine season (DV1, $ShotsPerPatientPop_{it}$)
    - Percent growth in 2018 was calculated for each clinic as the cumulative number of flu shots given divided by the final total of shots given in 2017 minus 1 (DV2, $PercentGrowth_{it}$)
    - Weekly change in shots per clinic patient was calculated for each clinic as the number of shots given that week divided by the total number of patients a clinic saw that vaccine season (DV3, $SPP\Delta_{it}$)
    - Weekly change in percent of clinic capacity was calculated for each clinic as the number of clinic shots given that week divided by the final total of clinic shots given that season (DV4, $PCT\Delta_{it}$)
- The following independent variables were constructed as follows:
    - For 2017 (all CC clinics) and 2018 (the CC Ranking clinics), the HighRank flag was set to 1 if that clinic ranked in the top 10% (rounded up) that week, and zero otherwise (Variable: $HighRank_{it}$)
    - For 2017 (all CC clinics) and 2018 (the CC Ranking clinics), the LowRank flag was set to 1 if that clinic ranked in the bottom 10% (rounded up) that week, and zero otherwise (Variable: $LowRank_{it}$)
    - The 10% definitions were relaxed and constricted to test robustness as discussed in the paper

The regression models were evaluated exactly as discussed in the paper using the above variable definitions and constructs.