The Contingent Effect of Management Practices

Steven Blader
New York University

Claudine Gartenberg
New York University

Andrea Prat
Columbia University

August 25, 2014*

Abstract

This paper investigates how changes in managerial practices interact with interventions on workplace values. A large US logistics company is in the process of fitting its trucks with an electronic on-board recorder (EOBR), which provide drivers with information on their driving performance. In this setting, a natural question is whether the optimal managerial practice consists of: (1) Letting each driver know his or her individual performance only; or (2) Also providing drivers with information about their raking with respect to other drivers. The company is also rolling out the first phase of a multi-year "lean-management journey". This phase focuses exclusively on changing employee values, mainly toward a greater emphasis on team work. The main result of our randomized experiment is that (2) leads to better performance than (1) in a particular site if and only if the site has not yet received the value-centered intervention. The result is consistent with the presence of a conflict between competition-based managerial practices and cooperation-based workplace values. More broadly, the experiment shows that the usefulness of a specific managerial practice may hinge on the value system that the company has put in place.

*We thank...

1 Introduction

Economists have increasingly focused on management practices as an important explanation for the large observed variation in productivity among firms (cit). Differences in practices, both across countries and industries (cit) and over time within individual firms (cit), have been shown to explain differences in productivity [replace with actual data from papers]. A logical puzzle, however, naturally follows from these findings: if there is such clear returns to management practices, why do we not observe firms either adopting them or exiting? Instead of a universal convergence of "state of the art" practices, we observe a wide variation in the rates of adoption among firms - presumably in equilibrium - [dimension "large?"]. [sentence on scale here.]

One possible explanation for this puzzle is that weak institutions add costs to the adoption of otherwise productive practices, leading to widespread non-adoption and survival of inefficient firms, particularly in developing countries (Bloom et al, 2013). In this view, managers might have either insufficient incentives, information or time to adopt these measures. Instead of being driven out of the market by competition, inefficient firms run by these managers survive because weak legal infrastructure constrains both incumbent expansion and new firm entry. While providing a convincing argument for firms in developing countries, however, this argument does not fully explain the patterns we observe, as we see a large variance in practice adoption in richer settings. Even more importantly, it also leaves open the question of why the within-country variance is XX greater than between-country variance, even when comparing countries with both strong and weak institutions.

A distinct but complementary explanation, derived from management research (cites) [bring up Milgrom-Roberts here], is that the large variance in practice adoption is, in fact, efficient and a manifestation of complex interrelationships between practices. That is, the impact of following a certain practice is contingent on the other practices implemented within a firm, some of which may enhance the effects of new practice and others of which may degrade them. Because of these complementarities and the cost of changing certain practices, particularly those that are "soft" and require longer periods to establish credibly (cite here?, FN examples?], different bundles of practices

will exist in equilibrium. This argument is related to Milgrom and Roberts (XXXX), who similarly propose complementarities between firm choices. If 'soft' practices are harder to observe and they vary from firm to firm for exogenous reasons,

Both explanations have been considered theoretically and using case examples (cit siggelkow, porter, rivkin, etc). However, while there is empirical evidence for the institutional argument, to our knowledge, there is a near complete absence of evidence for the contingency argument. The lack of empirical evidence could be an indication either that it is not a strong factor or of the measurement challenge associated with observing practices distinctly and manipulating them in such a way to isolate credibly any contingent effects.

One of the goals of our study is to demonstrate that the effects of management practices are, in fact, contingent on each other. We run a field experiment within a trucking company with a large number of establishments (which we refer to throughout as "sites") in a developed market context (the United States). For the experiment, we randomly assign the posting of individual performance metrics and rank - a formal workplace practice - stratified across sites by whether each had implemented a second, "soft" practice prior to the beginning of the study. This second practice is a program to alter the local culture of each site from an individualistic, low worker-manager trust culture to a Japanese-inspired team-focused, "servant leadership" environment [rephrase?]. At the time of our study, this program was in the middle of a quasi-random (for our purposes) rollout across the company.

Consistent with Milgrom and Roberts (XXXX), the effect of the performance posting treatment was strongly contingent on whether the sites had initiated the team culture program. Drivers assigned to sites with the individualistic values responded on average positively to the performance postings, improving their fuel efficiency by X% and reducing their idling time and wasted fuel respectively by Y% and Z% relative to the control group. In contrast, drivers at sites with the new team-based values responded negatively to the individual performance rankings. We record a substantial drop in performance for these drivers, in the form of an X% reduction in fuel efficiency and an increase of Y% in idling time and Z% in wasted fuel, relative to the control.

These results are consistent with both the contingent nature of management practices as well as

predictions from social psychology (cite) [two sentence here on these predictions - Steve? Here's my version, which should be replaced:] Workers in settings in which group identity is emphasized over the individual are likelier to perceive public postings of individual performance as pitting "team member against team member" and as a violation of the organization's values by managers. In contrast, workers in settings that stress individual identity and performance are likelier to respond positively to rankings and other public recognition of personal performance as consistent with managers' rewarding strong employees.

The results also provide support for a contingency explanation of why we observe such large variations in management practices within countries, even within those with strong institutions. Extending the example from our field experiment, if rank postings leads to high individual motivation, the practice may be adopted widely in individualistic firms and rarely in team-oriented firms. If the combination of rank postings and individualistic values has similar productivity effects as the combination of no postings and team values, then we may observe both combinations in equilibrium.

The advantage of our setting is that the organization of our firm, with a relatively large number of comparable sites and employees and constant underlying firm-level attributes, permits us to design a randomized experiment with statistical power and without the concern that unobserved firm-level factors are driving our results. Also, our study is less vulnerable to critiques about the external validity of a one-firm experiment since we are primarily concerned with demonstrating the existence of contingent effects between any combination of management pratices, rather than the specific ones used in the experiment. In other words, unless we believe that our setting is the only context with complementarities between management practices, then demonstrating their existence in our setting is sufficient to demonstrate their existence in general.

The biggest challenge in our experimental design is the assignment of the values intervention across sites. Ideally, both the rank posting and the values intervention would be randomly assigned. However, the constraints of our study required us to use a quasi-random assignment for the values intervention: the posting treatment had a specific three month window to be rolled out across all the sites, while the values intervention required a minimum of three to six months per site and

substantial management focus to take effect. Given this timing mismatch, we chose to stratify across sites based on whether they had received the values intervention at least three months prior to the commencement of the study and further applied statistical matching between firms with and without the intervention. Noteably, our results generally strengthened or remained stable after the matching.

It is possible that this empirical challenge is unavoidable, given our research question. If all management practices could be costlessly and immediately implemented - like our performance postings (approximately) - then we might expect much more convergence among practices. Although we do not show this in our study, we speculate that it might be complementarities between "soft" practices that are costly to change - such as the values program by our trucking company - that leads to the high dispersion even among "hard" practices that are easy to measure and implement.

This study has several implications. The most direct one is that, by showing that practices are not simply additive but can have complex interactions, it proposes a different explanation for large differences in adoption rates of management practices. (repetitive?). [followon sentence about lit?] It also shows that "soft" factors can have a significant impact on the implementation of other practices and, by extension, on the productivity of organizations. [remove or replace next sentence about why this matters?] To the extent that we can interpret these soft factors a "relational contracts" in the sense of Gibbons and Henderson (XXXX), this study provides evidence that these relational contracts can in fact have a first-order impact on worker behavior and productivity.

Third, the study raises a methodological question about the generalizablity of findings from field experiments, a rapidly expanding technique within economics (cite). Specifically, studies that claim the direction and magnitude of an effect, but do not consider possible underobserved site-specific interactions, may draw fundamentally incorrect conclusions about the applicability of the findings. Using our experiment as an example, had we run our study in 2010, prior to the commencement of the values initiative, and omitted any consideration of the values of each site, we would have concluded that performance postings have a positive impact on productivity. Alternately, had we run our experiment in 2018, after the completion of the values initiative, we would have concluded the opposite. In this sense, our study sounds a cautionary note about the extent to which findings

from field experiments can be extended beyond their settings.

Lastly, this study has implications for research on the effects of employee rank postings (cites). To date, ranking studies have found mixed results, with some findings a positive impact on performance (cite) and others a negative or neutral result (cite). Our research provides one candidate explanation why these studies - which generally take place within single firms - have found inconsistent results.

2 Literature Review (To Do)

3 The Research Setting

3.1 Why the Trucking Industry?

The US trucking industry as several features that make it well suited for research on management practices. Intense product market competition and well developed information markets (in the form of trade organizations, conferences and consultants) leads firms to rapid adoption of productivity-enhancing technology. A subset of technology in the process of widespread adoption at the time of our study enabled managers to implement a broad range of previously infeasible driver monitoring and incentive practices.

Of particular interest for research, these managers then face (tense - present or past) the challenge of selecting and implementing the set of practices using the new technology that will most improve driver behavior. According to our interviews with managers, these practices are central to the success of these initiatives since technology itself.can be implemented in a wide range of configurations without consensus on a universal "best" way. In a worse case scenario, if implemented poorly, drivers can respond by collectively opposing the initiative and sabotahing the technology (FN: visors down, terminals turned off, bad calibration, etc.) This was a common concern among the interviewed managers, who viewed failed implementations as not only a significant waste of time and money but also as harming delicate manager-driver relations.

One of these technologies was being rolled out at our research site at the time of our study and

was of particular importance to our research design. Electronic on-board recorder systems (EOBR) (cite Baker Hubbard here?) record and transmit detailed driving behavior to a centralized database accessible to managers. This database can be used to evaluate, discipline and reward drivers in near-real time. EOBR systems also include terminals installed in truck cabs that display driver performance information and emit audible real-time alarms when driving behavior is determined to be inefficient.

Both of these features can be perceived by drivers as extremely intrusive, particularly given the independent nature of the work prior to EOBR introduction and the resulting types of employees attracted to a driving career (FN: the common view of the trucker as "the last American cowboy," with a "sense of fierce independence, counter-cultural defiance, and unapologetic masculinity that defined trucker culture." (FN: http://freakonomics.com/2009/02/27/ask-an-economist/). As a consequence, managers at our research site and other trucking firms were especially concerned with implementing the technology such that drivers viewed the new system as a benefit to them or, at minimum, not a threat. According to our interviews, this perception depended on how managers used the newly gathered data - in other words, the practices accompanying the technology.

Finally, the trends in trucking are occuring across many other industries and both low and high skilled job types (Brynjolfsson and other cites?). This raises a biggest question of job design as sophisticated monitoring technology is widely rolled out. (cites, more here)

3.2 The Company

The company at which we conducted this study operates in the less-than-truckload segment of the industry, transporting shipments that are smaller than full truckload freight and larger than individual parcels. The company employs approximately 10,000 drivers, all of whom are non-union hourly employees, across approximately 300 sites in the US and Canada (FN: numbers fudged a bit to preserve anonymity). Important for our study is that most drivers operate local routes and there is little communication between sites. Shipments are picked up and delivered during regular business hours via local routes of less than 200 (300?) miles covered by drivers that generally remain with the same customers over months or years. Intercity shipments are transported between sites

via by a minority of drivers, typically on an overnight shift. Because of the difference in shift schedules and the small proportion of line-haul drivers, the threat of cross-contamination between sites during our study is limited. This features enables us to establish a credible control group for our study.

The company was engaged in two major initiatives at the time of our study that we incorporated into our research design [rephrase so it doesn't seem to opportunistic - just that we combined three iniatives for our study]]. First, beginning on August 21, 2014 and continuing over a four month period, EOBR and safety cameras were rolled out to all trucks. [add: from a base of no monitoring to state-of-the-art - name the features added, also evidence/ quotes about paranoia by drivers about how data would be used]

Second, beginning in 2011 and continuing during our study period, the company was engaged in a five- to seven-year program to change their business culture and practices to conform to Japanese manufacturing practices. At the time of our study, the company had initiated the first of a five phase transition to this "lean" operations model, with a plan to complete the first phase across the remaining sites in 2014 and 2015. The focus of the first phase is primarily on transitioning the culture of the site from an individualistic and hierarchical culture to a team-based one in which the front-line workers are given substantially more [finish this sentence]. [one more sentence on japanese concept of team values, based on Toyota practices of team production translated to this context, drivers are responsible for local practices to improve performance and solving local problems,etc. The role of management is to provide firm support and resources to enable the changes. In contrast, the old way is that the drivers clock in and out, and all changes are driven centrally. More color here - "hard as rock" quote?]

The timing of this initiative had two advantages for our study. First, the first phase of this transition was primarily focused on enacting a team-based culture and involved no changes to formal driver-related workplace practices that could otherwise affect our performance measures (FN: outside of dock cleanup, which does not impact our measures). The second through fifth phases did focus on the formal tools side of lean manufacturing, but importantly, none of these phases had been initiated at the time of our study.

Second, approximately 1/X of the sites had begun this cultural initiative at least three months prior to the beginning of the study, enabling a meaningful comparison between sites that had undergone the initiative and sites without any culture shift.

Third, after the initial pilot phase, the rollout was unrelated to the anticipated success of the initiative. For that reason, we consider the rollout to be quasi-random for our purposes, in the sense that the rollout schedule is orthogonal to the unobserved characteristics of a site that may influence the success of the initaitive (or of the ranking posting - add this). (what about SACs).

Appendix XX shows examples of the criteria used to assess sites at the end of the first phase. Criteria such as "XX" and "YY" show the degree to which "soft" behavioral changes, rather than hard practices, are the focus of this phase. Appendix XX provides excerpts from interviews with supervisors and drivers about the effect of this initiative on the culture of the sites and employee behavior. [follow up sentence here]

4 The Experiment

[note somewhere that no formal incentives are attached to experiment or to drivers on the whole.] The experiment occurred between August 2013 and July 2014 as EOBR was rolled out throughout the company. We adopted a three by two research design where we assigned three ranking conditions randomly across sites for which collective and individualistic values were quasi-randomly distributed.

4.1 The Ranking Intervention

For the ranking conditions, aside from the control group, we designed two ranking treatments: one in which the driver names were posted next to the rankings and one in which the employee IDs were used. In this latter treatment, a driver can identify his own standing and view the distribution within the site, but does not know any other individual's rank, nor do others know his rank. [sentence here on why we decided to have these two conditions] Because of the substantial number of sites and the lack of pre-existing outcome data to to perform power analyses, we placed equal numbers of sites to each of the three conditions.

The rankings were posted on a weekly basis, beginning six weeks from the EOBR rollout date for a given site. This timing allowed us to obtain thirty days of pre-measures (we discarded the first two weeks while the systems were calibrated to the trucks). The pre-measures, combined with the control group and lean stratification, enable a triple differences research design that we describe section XX.

The rankings contained either the driver names or their employee IDs, depending on the treatment assignment, and four performance metrics as reported by the EOBR system. These metrics are gap score, shift score, excess idle time and total fuel lost. We discuss each of these in more depth below in Section XX. See Appendix XX for pictures of the postings. The ranks were sorted based on the gap score (FN: not consistent compliance with the sorting and display of rankings), but other performance metrics were also clearly visible. [sentence on lack of standardization of postings? Is needed?]

4.2 Team Values Initiative

For an ideal experiment, lean status would be randomly assigned, similar to the ranking conditions. However, this assignment was impractical given the constraints of the study timing and the long timeframe and high management intensity required to alter site culture. The study was constrained to coincide with the EOBR rollout during the last four months of 2014 since rank postings were bundled into the larger EOBR rollout initiative by the company. In contrast, altering site culture from an individualistic focus to a team focus requires at minimum three to six months per site and could not be implemented on a large enough scale to be integrated with the study. (FN: see BGP 2015 for a study that does randomize this intervention). Instead, we stratified our ranking assignments by lean rollout status as of the start date of the study and applied statistical matching techniques to correct for differences between lean and non-lean sites.[concluding sentence here - reread bloom and all for level of discussion]

4.3 Sites Included in Field Experiment

[doublecheck for anonymity]

The firm has a total of approximately 275 sites. Of these sites, we discarded 25 from the sample that launched the cultural change initiative in 2011, the launch year of the program. In discussions with the company senior management on the need for quasi-random assignment of lean, they mentioned that the earliest sites were selected for either exceptional leadership or workforce or other conditions directly related to the anticipated success of the rollout within those sites. While the company selected less than ten sites [doublecheck] for these initial pilots, we conservatively discarded all sites launched in that first year. We further excluded 36 sites that were scheduled to launch the values initiative during the timeframe of the study, as these locations could not be reasonably be classified as having individualistic or collective values. Lastly, we discarded 72 of the sites with rankings launched in the first half of the study (prior to November 25, 2013). Until the mid-project checkpoint in early November, no formal verification process of the rank postings had been instituted and upon further investigation, we determined that there was widespread noncompliance up to that point. Beginning November 25, the company instituted a formal process to verify that ranks were posted as required by the experiment guidelines, include weekly photographs of the postings, conference calls, and a shared spreadsheet tracking system.

After these corrections, the experimental sample included 142 sites, 47 in the control group, 50 in the named postings group ("Treatment 1") and 45 in the IDed postings group ("Treatment 2"). To construct our driver-day sample from the EOBR system, we then further removed line-haul routes (defined as routes above XXX miles) and routes with EOBR data that was clearly unreasonable (FN: less than 15 mile routes or MPG<1 or >15). This left us with a sample of 330,689 driver-days.

Because the company did not have the managerial bandwidth to reinforce the ranking messaging on an on-going basis, we expected to see some diminishment of any effects of the postings over time. For our main multivariate analyses, therefore, we restricted the windows of the experiment from the thirty days prior to the postings to the thirty days after. We also removed the five days immediately surrounding the scheduled posting dates, since many of the site managers chose a different day of the work week to post the rankings to coincide with group meetings, rather than on the date specified by the experiment. We were left with a sample of 93,913 driver-days within these narrowed windows that we use in our primary multivariate analyses, although in the appendix we repeat the analyses

with the long windows and show that most of the results are replicated, although with somewhat larger standard errors.

Appendix Table A1 contains a summary of the sample construction process.

4.4 Outcome Variables

We focus on four outcome variables for this study that capture different aspects of efficient driving performance. Gap score, the primary measure used in the driver rankings, calculates the difference between the average actual and "potential" miles per gallon expended on a given route. The potential miles per gallon is calculated based on what the EOBR system considers to be optimal shifting and speed patterns, given weather conditions and route characteristics. Gap score is represented in percentage terms such that, for example, if actual and potental mpgs for a given route are 7.0 and 6.5, respectively, the gap score would be 7.7 ((7.0-6.5)/6.5). A higher gap score, therefore, represents worse (less efficient) driver performance.

Shift score is the percent of shifting events performed on the route that remains within designated RPM limits for the engine. For example, if a driver shifts five hundred times on a given route, his shift score will be 90 if he revs the engine above a designated threshhold during fifty of those shift events. In contrast to gap score (and the two other measures described below), a highe shift score represents better driver performance [### CBT: should we reverse-score shift score and redo tables????]

Excess idle time is a measure of the minutes an engine idles for longer than a designated time period (doublecheck) and therefore wastes fuel. Lastly, fuel lost is an aggregate measure of all the fuel wasted from idling, inefficient shifting, speeding and gearing. For both of these two measures, a higher value represents worse performance.

Because all four measures are included in the weekly rank postings, we investigate each of them as outcome variables in our analysis.

Table 1 reports summary statistics for the sites in the field experiment. Note that the sample of 142 sites used in the experiment is generally representative of the full set of 275 sites within the firm, based on observable site characteristics and pre-rank posting driver performance. Within

the sample, the Control and Treatment 1 (named ranks) groups are statistically indistinguishable, while the Treatment 2 (IDed ranks) group appears somewhat difference from the other two groups, particularly in the distribution of lean sites and pre-posting driver performance. We both control explicitly for this bias in our primary analysis and perform additional analyses using a matched sample that removes the biases. However, even with these corrections, we interpret results concerning the Treatment 2 group with more caution than those of the Treatment 1 group.

5 Impact of Rankings and Collective Values on Driver Performance

5.1 Preliminary evidence

Figure 1 depicts the combined response of both lean and non-lean sites to the driver postings, as measured by the each of the four outcome variables. The x-axis is normalized such that week 0 represents the week that rankings were posted at each site, regardless of the calendar date of each posting. The y-axis measures the outcome, with higher values generally signifying worse performance (with the exception of shift score where higher values signify better shifting behavior). We aggregate the data in two complementary ways. In Figure 1a, we plot the median values of each outcome and weight each site equally, regardless of the number of drivers per location. In Figure 1b, we plot the average values of each outcome variable across all drivers, effectively weighting each driver equally regardless of site assignment.

Four results can be drawn from this figure. First, within the control groups, the performance outcomes either deteriorate or remain relative constant over time after the EOBR system is launched (at week -6). For example, in Figure 1a, gap scores for the control group range between 1.2 and 1.35% for weeks -6 to 0, and increased to 1.3 to 1.49% for weeks 1 to 18. A similar pattern is apparent for both shift score and fuel lost. According to our field interviews, this trend may be attributable in part to the reluctance of management to assign any formal incentives to driving performance, which may in turn have led to reduced focus on the EOBR system over time.

Second, the pre-posting trends appear similar between between the named rankings (Treatment 1) group and the control group, consistent with successful random assignment of sites between groups before the start of the experiment. Third, after rank posting at week 0, there is a clear divergence in Figure 1a between the named rankings group and the control group, with the exception of the fuel lost metric. This pattern is not apparent in Figure 1b, which plots the outcomes as averages across all drivers. The difference between Figures 1a and b can be explained by the greater number but smaller sizes of the non-lean sites relative to the lean sites. Figure 1a, therefore, places more weight on the response of non-lean sites to the treatments and show that for these sites, the ranking treatment had an overall positive effect on driver performance. We explore this result further in Figure 2.

The final result apparent from Figure 1 is that the IDed rank group (Treatment 2) appears to be different from the two other groups in pre-measures of gap score. Although these differences are not statistically significant (see Table 1 - and check to see whether fig1b pre-posting averages match table 1), we would have preferred to see closer pre-trends between the two groups. Because of these potential differences and as discussed in Section XX, we focus more on the named ranking group for our analyses and we draw conclusions from the IDed rank group with more caution.

<< Insert Figure 1 about here >>

Figures 2a and b depict the performance response to the ranking postings by individualistic and collective (standardize terminology here) sites and show graphically one of the main results of the experiment. For visual clarity, we replace the absolute levels of the performance metrics as shown in Figure 1 with the differences between the treatment and control groups. As such, the 0-line on the Y-axis represents no difference between the two groups, a positive difference (shown by a line above the 0-line) represents worse performance for the treatment group and, conversely, a negative difference (shown by a line below the 0-line) represents better performance (note that, for shift score the opposite is true).

Three results are apparent from this figure. First, each of these four plots shows a clear separation in response between lean and non-lean sites to named postings, with the non-lean sites in the

named rank group showing relatively better performance than the control group (above the 0-line) and the lean sites showing relatively worse performance (below the 0-line).

Second, the difference between the control and named ranking treatment group appears somewhat persistent over the timeframe of the experiment, although near the final weeks of each plot, the performance measures converge somewhat. Regarding this convergence, several points are important to note: there is more noise in these final weeks than in the earlier weeks, since only the sites with the earliest ranking rollouts had data that extended this far at the time of the analysis Also, the company did not have the managerial capacity to reinforce the salience of the ranking postings consistently throughout the study period, so some reversion was expected. As a result, it is not possible to infer whether apparent convergence is an artifact of the data and experiment or is a more general finding.

<< Insert Figure 2 about here >>

Overall, this graph shows preliminary evidence for one of the main results of the experiment: that driver reponse to social incentives is contingent on the communicated values (terminology here) of the site to which he is assigned.

5.2 Intent to Treat Estimates

We now turn turn to multivariate analyses to verify the preliminary results discussed above. We estimate the differential impact of the rankings on lean and non-lean sites using the following triple-differences equation:

$$PERF_{it} = \alpha TREAT_GROUP_i * LEAN_i * POST_{it} + D'_{it}\beta + e_{it}$$

where i represents a given driver and t is the calendar date. $PERF$ is one of the four performance outcomes, $TREAT_GROUP$ is a vector of two indicator variables, one for each of two possible treatments assignments (named or IDed rank postings), and $LEAN$ is an indicator variable that is equal to one if the sites have had the collective values initiative introduced at least three months before the beginning of the experiment. $POST$ is equal to one after the assigned date of the ranking rollout for the two treatment groups or, equivalently, six weeks after the EOBR rollout for the control group. All pair-wise interactions and individual variables associated with the triple-

differences term are also included in the model and α represents the vector of coefficient estimates for all the associated terms. We are primarily interested in the coefficient on the triple interaction itself, which estimates the difference in response to the rank postings of lean and non-lean sites. D_{it} is a vector of control variables that includes site characteristics, day of week indicators to absorb weekly patterns, lean manager FE, etc (rephrase).

We also perform several variations of this analysis. We add in driver and date fixed effects to control both for driver traits and seasonality. We also create a subsample of the data that matches lean and non-lean sites to account for the non-random assignment of lean in the experiment. Lastly, we do an instrumental variables analysis where we instrument X with Y [fill in] to account for incomplete compliance. We discuss each of these analyses in turn below.

5.2.1 Combined Effect Across All Sites

We begin by estimating the simple intent-to-treat model without differentiating between lean and non-lean sites. Table 2 shows the results of all four performance outcomes, both without and with controls (odd and even columns, respectively). Consistent with Figure 1b, the ranking intervention appears to have no effect. Without accounting for the undrelying values at the site, therefore, we might inaccurately conclude that social (rank) incentives have no effect on worker performance.

<< Insert Table 2 about here >>

5.2.2 Effect by Lean and Non-Lean Sites

We next estimate the differential impact between lean and non-lean sites Table 3 shows the results of the intent-to-treat analysis modeled in Equation X and is consistent with the plots in Figure 2. In Columns (1) and (2), we see a large, positive difference in the response of lean sites to the named rank treatment. An increase of 0.1358 in the log value of gap score represents 22% of a standard deviation of gap score. Non-lean sites see a reduction in gap score of 0.0423 (logged and non-logged), representing 7% of a standard deviation improvement in performance.

No similar effect was estimated for the second treatment group that posted the IDed ranks. Consistent with successful randomization, we estimate no statistical difference between treatment

groups and control groups (coefficients on Treatment Group 1 and 2 and Lean X Treatment Group 1 and Lean X Treatment Group 2), nor any difference between the control groups in the post-rank posting response (coefficient on Post X Lean) We do, however, see some evidence of overall performance deterioration for the control group (coefficient on Post), possibly reflecting the earlier observation by company management that the lack of explicit performance incentives may have led to decreased attention paid to the EOBR system over time. Similar results are evident in Columns (3)-(8) for the other three performance outcomes.

Interestingly, the deterioration in the driving performance within lean sites in response to named rank posting is consistently more significant and larger than the improvement in the non-lean sites. For example, the reduction is shift score of 0.845 points (10% of a standard deviation) from the Column (3) estimate is larger than the 0.0968 point improvement (1% of a standard deviation) estimated for the non-lean sites. [comment here on how to interpret this result and how to square with aggregate results (# sites and drivers in each treatment, etc)].

Also notable here and in subsequent analyses is that the inclusion of control variables does not change the estimates substantially, consistent with a successfully randomized experiment.

[also add section for calculating economic magnitude:] Using the coefficient on the fuel lost metric in Column (8), we try to further quantify the economic magnitude of this effect. An increase in 0.0553 gallons lost (per mile?? - do we need to multiply by # miles to get total fuel lost? email Ryan]. << do dollar analysis here>>

<< Insert Table 3 about here >>

Table 4 repeats the analysis including date and driver fixed effects and the results remain largely unchanged. The ITT estimate in Column (1) of the Treatment Group 1 indicator on Lean sites of 0.1026 represents 26% of the within standard deviation for log gap score on drivers, similar to the estimate in Table 3. Also like Table 3, we observe no underlying pattern for Treatment Group 2, the IDed rank group. The similarity in results between Tables 3 and 4 supports the conclusion that the effect of ranking is due to changes in driver behavior and not due to compositional differences between sites.

<< Insert Table 4 about here >>

5.3 Matched analysis

One challenge of the study is that lean status is not randomly assigned. A partial mitigation to this concern is that the company's management indicated to us that - after the 2011 pilot period (not included our sample) - the choice to launch the values initiative at a given location was driven by geographic expediency and schedule optimization for the managers involved in the initiative. However, as is evident from statistics in Table 5, this rollout strategy still resulted in underlying differences between lean and non-lean sites, with the average site size as the biggest difference between the two groups. For our experiment, this size difference poses a challenge in interpreting our main result if, for example, large sites are both lean and more likely to resist rank postings for unrelated reasons. Table 5 also shows other underlying differences between sites, including lower MPG in lean sites, possibly reflecting more urban locations and therefore different route structures.

For this reason, we construct a matched subsample of sites that adjust for these underlying differences. This subsample includes 78 of the 142 sites in the full experimental sample. The excluded sites include nine of the larger sites in the lean group and fifty five of the smaller, more rural sites in the non-lean group. The right side of Table 5 compares the sites across these two groups in this reduced sample and shows that the sites are now statistically indistinguishable across both observable site characteristics and pre-rank posting driver performance.

<< Insert Table 5 about here >>

Appendix Table A2 reproduces the descriptive statistics in Table 2 using the matched sample and shows that the two treatment groups are now indistinguishable from the control group [doublecheck].

Table 6 reproduces the ITT estimates of Table 3 on the matched sample. The point estimates largely increase and reflect larger percentage standard deviation increases [doublecheck]. For example, the Column (2) estimate on Post*Treatment Group 1*Lean of 0.1510 represents a 25% standard deviation increase in log gap score, compared to 22% increase based on the estimates in Table 3 Column (2). In general, the results of Table 3 are reproduced and, if anything, strengthened, with the exception of shift score. [sentence to speculate on why?]

<< Insert Table 6 about here >>

Table 7 replicates the fixed effects analysis of Table 4 on the matched subsample. The results are largely reproduced, again with the exception of shift score.

<< Insert Table 7 about here >>

5.4 Additional Analyses

5.4.1 Potential Changes in Underlying Route Characteristics

We also performed three additional analyses to address separate challenges regarding data and experimental design. One question is whether there is anything fundamentally different about the structure or characteristics of the routes driven in lean and non-lean sites that could result in the same observed patterns over this time period. While we believe that this possibility is remote - it would have to affect only the named rank treatment group, only at the same six week post-EOBR rollout window as the rank postings and also be orthogonal to the site characteristics on which we based the matched analysis - we perform a placebo test to further rule out this possibility. For this test, we replace our four outcome variables with "potential" MPG. Potential MPG is the system-calculated variable that response to route characteristics and road and weather conditions, but not to driver performance. Therefore, if any route characteristics changed during this period in the lean named-rank group that led to changes in driver performance, we should observe similar patterns in the potential MPG metric.

The results of this analysis are shown in Appendix Tables A3 (on the full sample) and A4 (on the matched sample). Potential MPG shows no changes during this period, while Actual MPG - which is directly related to driver effort - does. It does not appear, therefore, that underlying changes to the routes are driving the results.

5.4.2 Correcting for Compliance

Coordinating the posting rollouts posed a management challenge for the company, particularly since the rankings were rolled out on a weekly basis across XXX sites in 48 states during the busy winter season. As a consequence, approximately 50% of the sites complied with the rank postings during

our experiment [do this by site-date? look at updated compliance spreadsheet]. To account for this selective compliance, we followed the approach from [CITE HERE] and instrumented actual treatment with assigned treatment. The results of this analysis are shown in Appendix Tables A5 and A6 and are stronger than our earlier analyses (including fixed effects and the matched cohorts). These stronger results are consistent with [choice to comply...]. Further, they provide evidence that our main results likely understate the economic impact of the rank postings on driver performance, particularly for XXX.

5.4.3 Persistence of effect

Finally, Appendix Tables A7-A10 repeat the analysis shown in Tables 3-6 without restricting the time windows to the 30 days pre- and post-rank postings. These analyses now include XX days prior and YY days after the rank postings. We find that, consistent with some attenuation, the standard errors of the estimates are generally larger, but the effect sizes remain essentially unchanged from the narrower-window analysis. [doublecheck]

[what do we want to do with undoing the rankings? - don't lose this idea]

5.5 Asymmetric Response Across Drivers

The ITT analysis estimates the average response of drivers to the rank posting within each treatment group and type of site. However CITE and CITE predicts (our theory above?) that worker utility (and, by extension, effort) differs based on the standing of that worker relative to her peers. To test this prediction, we examine driver response to rank postings based on their pre-rank performance. (Need Andrea input into this paragraph)

Figure 3 plots this analysis across the four outcome metrics. The x-axis represents the driver performance during the 30 days prior to the rank posting and the y-axis represents the driver performance during the 30 days after the rank posting. The 45-degree line in each plot, therefore, is a theoretical baseline for which driver performance remains unchanged between the two periods. The right and top of each plot shows worse performance in the pre- and post-periods, respectively, and the further below the 45-degree baseline, the greater the performance improvement between

the two periods (conversely, the further above the 45-degree line, the more performance degraded). [limit analysis to drivers with at least 20 days of driving during the analysis period - make sure to put this somewhere, maybe table footnote]

The plots in the left column show the combined driver distribution across lean and non-lean sites, breakout of the control and treatment groups separately. We expect some degree of mean reversion, which we can see from the fact that the slope of the control group is consistently less than 45 degrees and crosses the 45-degree line. However, beyond mean reversion, we see that the best-fit polynomials of Treatment Group 1 and 2 lie below the control group in all four plots primarily in the right-bottom quadrants of the plots, the region representing the worst drivers during the pre-period. From this result, we can infer than aggregate performance improvement - if any - is driven by the worst-performing drivers during the pre-period.

The middle and right hand columns shows the driver distributions for non-lean and lean sites, respectively. We find consistent differences in the distribution of driver responses across the two groups, particularly to the named rank treatment (Treatment 1) .Within the non-lean sites, the driver distribution within the named rank treatment group generally lies below the control group, with the distance between the groups increasing as driver pre-posting performance decreases (on the right of the plots). This increasing distance signifies that, within these sites, the worst drivers had the greatest performance improvement.

We see the opposite pattern for lean sites, the plots in the right-hand column of Figure 3. For these sites, the driver distributions of the two treatment groups consistently lie above the control group and closer to the 45 degree line. The distance between the control and treatment groups also increases to the right of the plots, although with the reverse pattern as with the non-lean sites: the slope of the best-fit polynomial of the treatment groups is consistently greater than that of the control group and closer to the 45-degree line. From this result, we can infer the opposite response within the lean sites: that the worst-performing drivers had the most performance degredation during the post-rank posting period. These patterns are replicated with site-demeaned data as well, as shown in Appendix Figure X [add this figure]

<< Insert Figure 3 about here >>

Table 8 shows these findings in multivariate form. [need to redo tables to have them as triple interaction between lean*pre-perf*treatment 1 - expect positive coefficient here]. The odd columns show the results for all drivers XX [explain what is in each column]. The positive coefficient estimate on the interaction between lean and pre-treatment performance indicates that ..[REDO THIS TABLE]

<< Insert Table 8 about here >>

6 Discussion

7 Conclusions

7.1 First reference

Figures

Figure 1: Impact of Rankings on Driver Performance

Figure 1a: Median effect with each site weighted equally

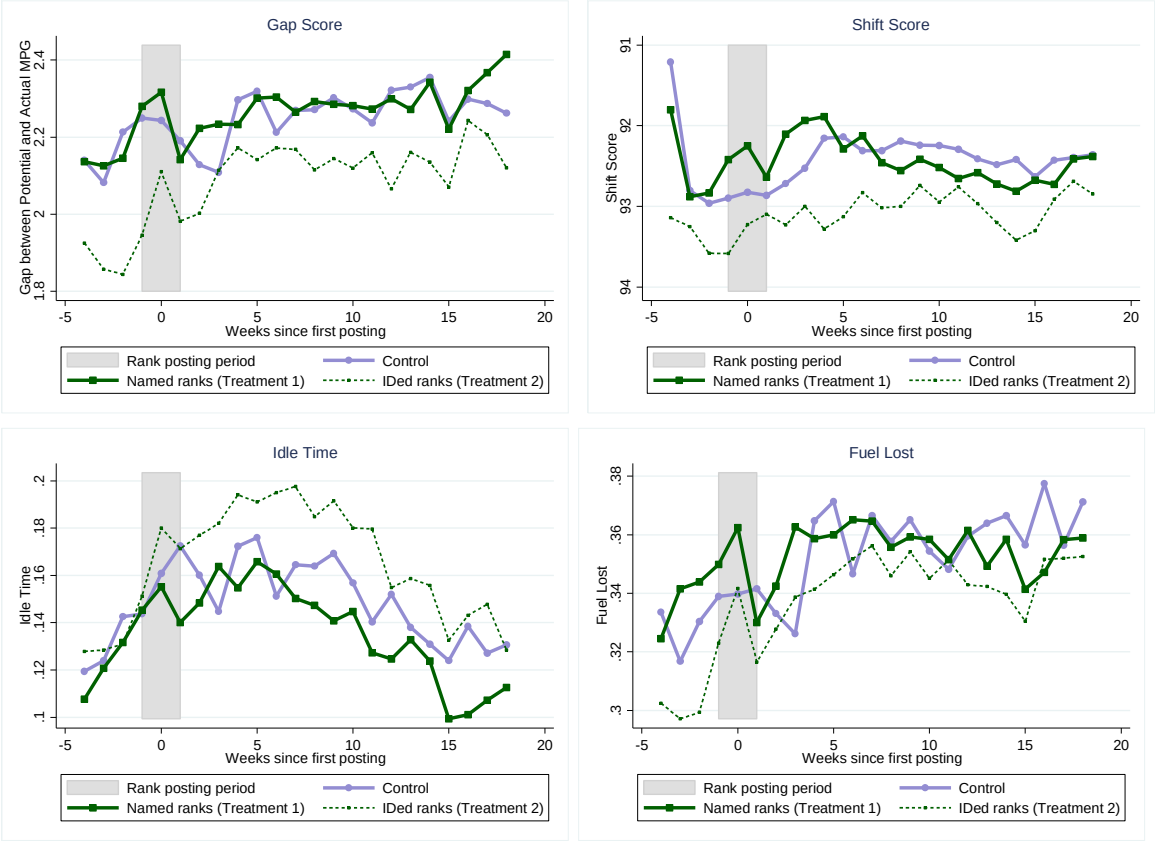


Figure 1b: Average affect with each driver weighted equally

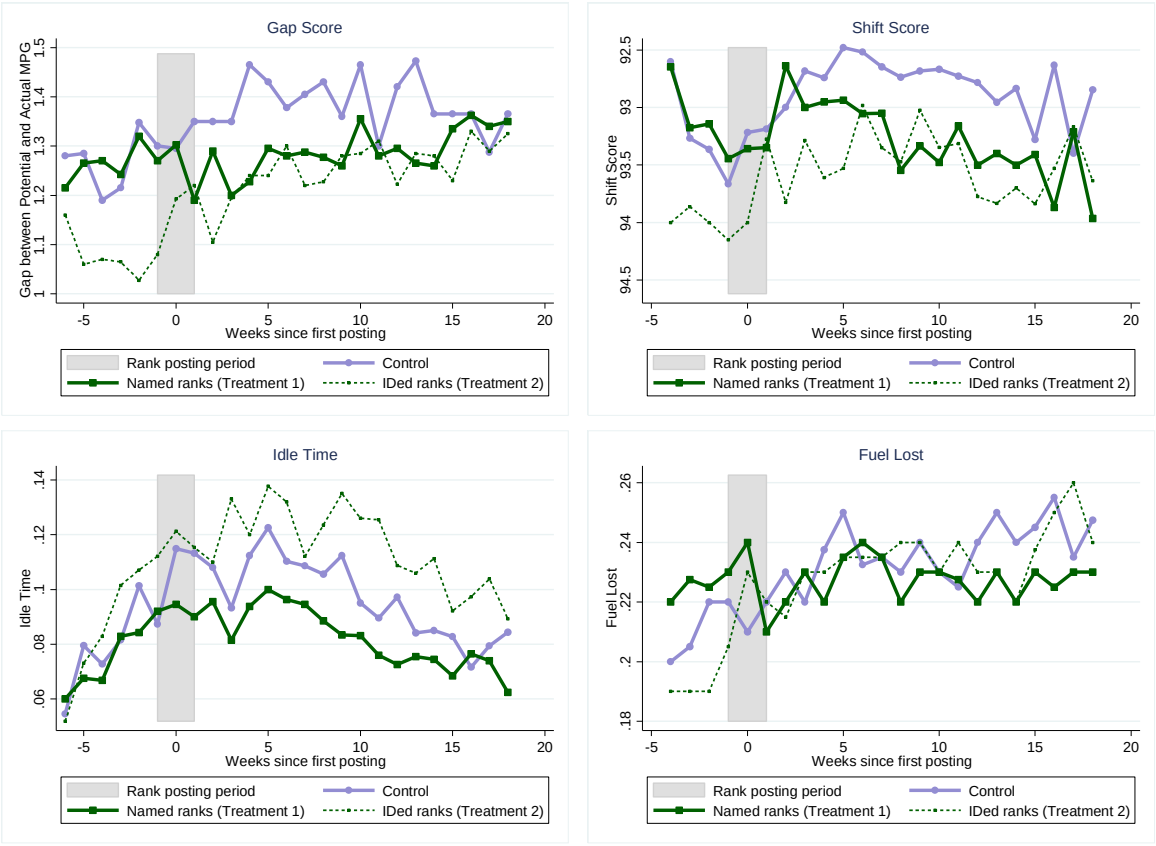


Figure 2: Impact of Rankings by Site Type

Figure 2a: Median effect with each site weighted equally

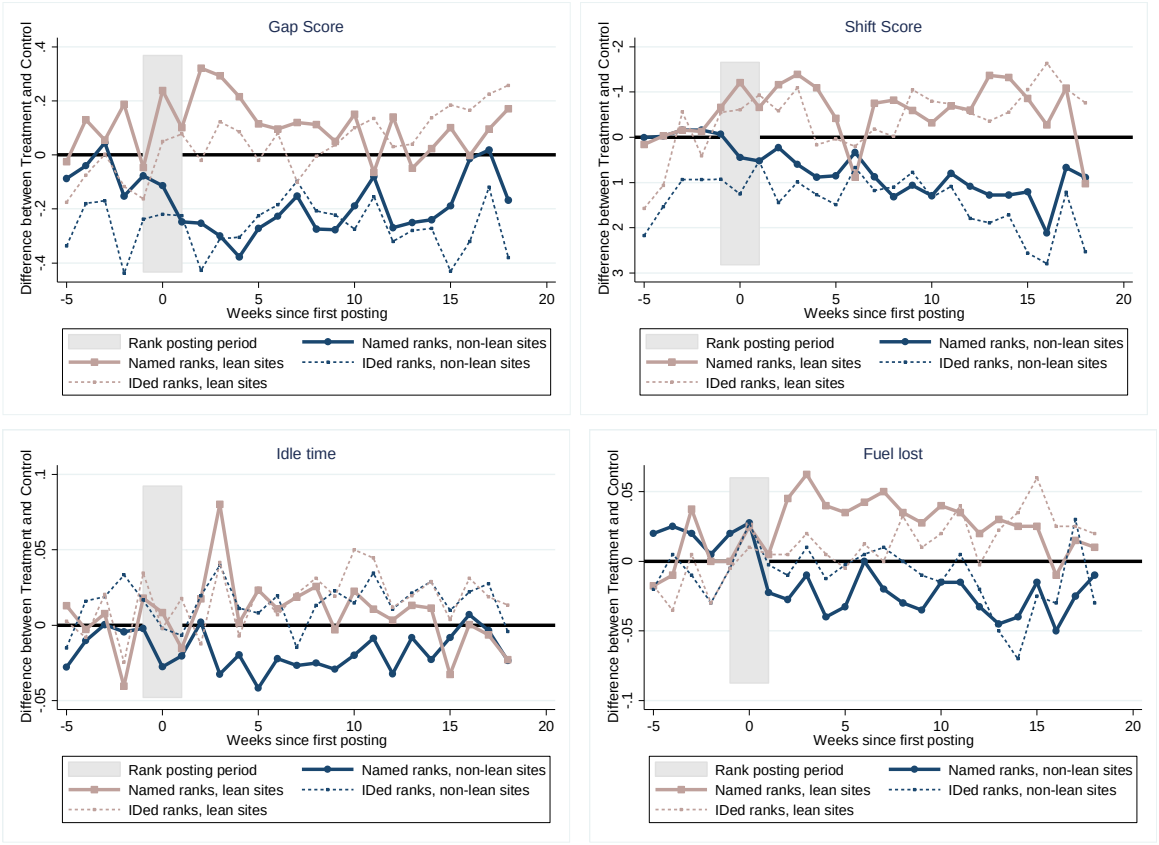


Figure 2b: Average affect with each driver weighted equally [doubleck old graphs – why did top end result go away?]

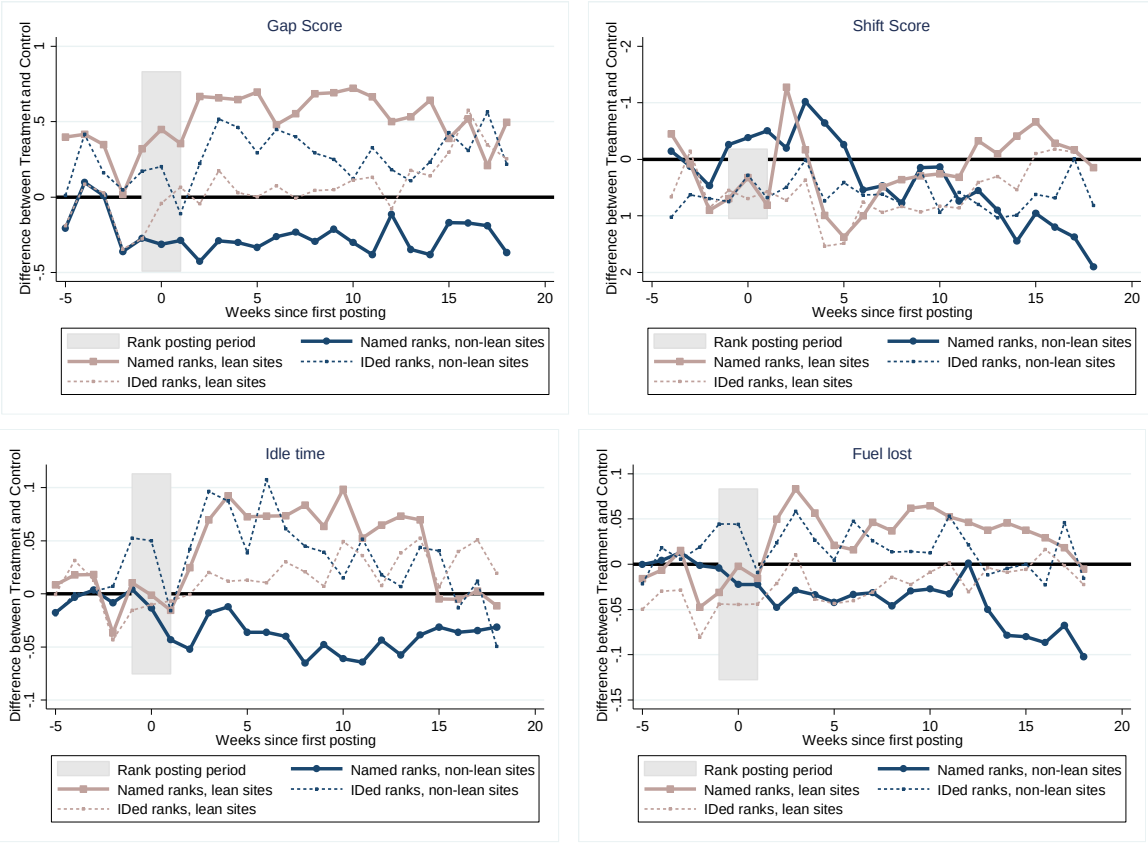


Figure 3: Asymmetric Treatment Response by Lean Status

Figure 3a: Gap Score

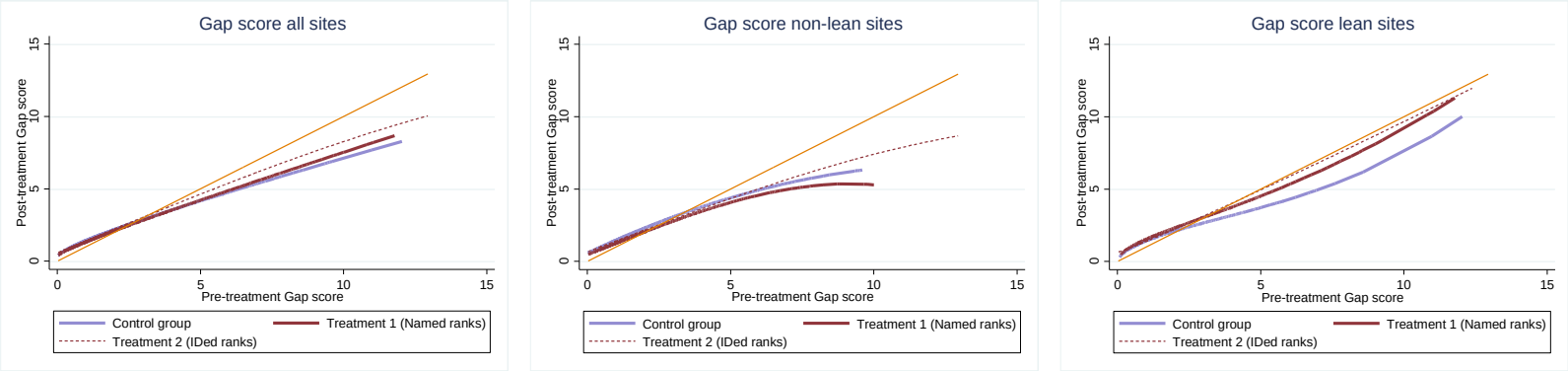


Figure 3b: Shift Score

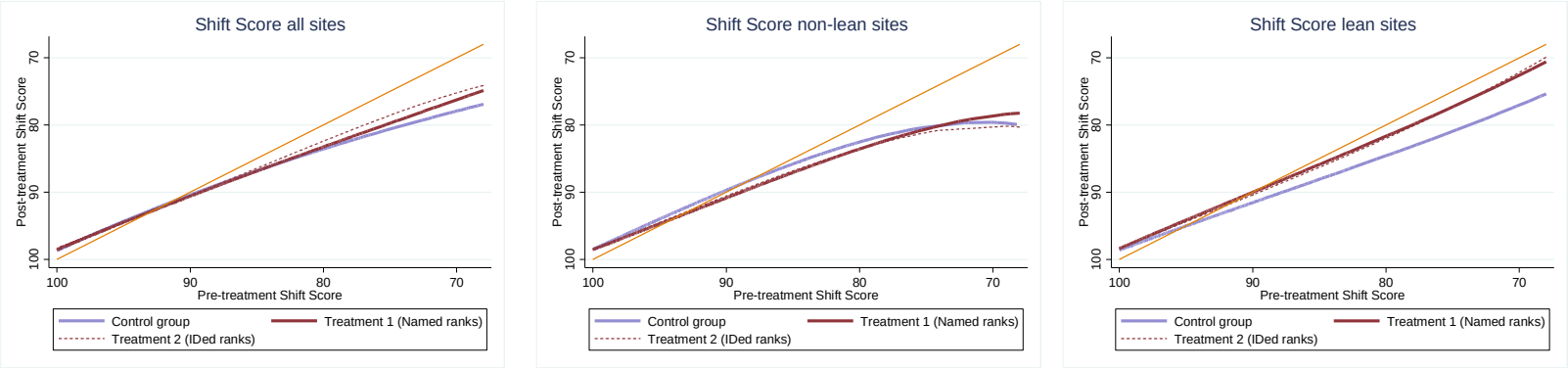
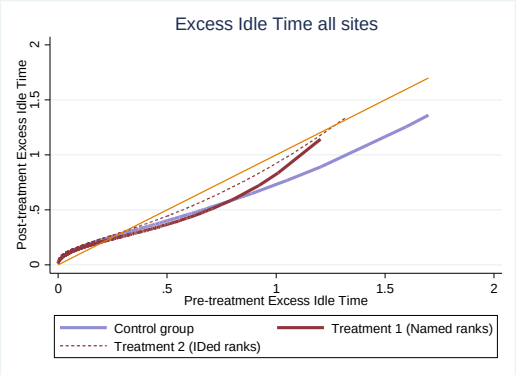
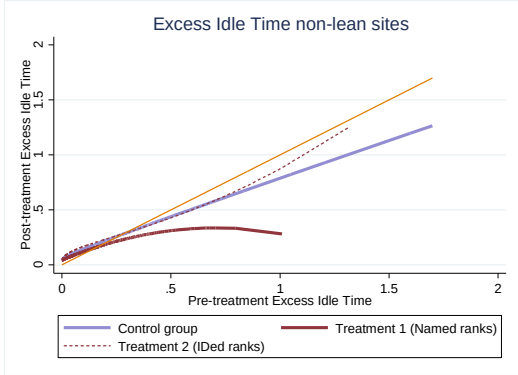


Figure 3c: Idle Time





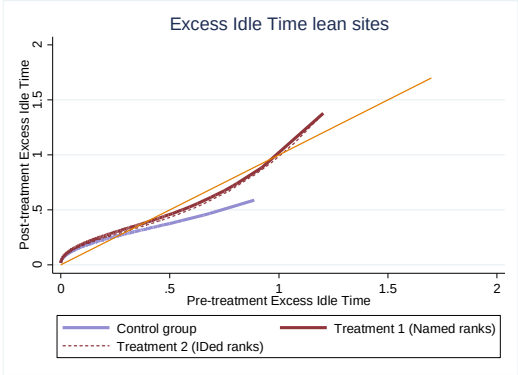
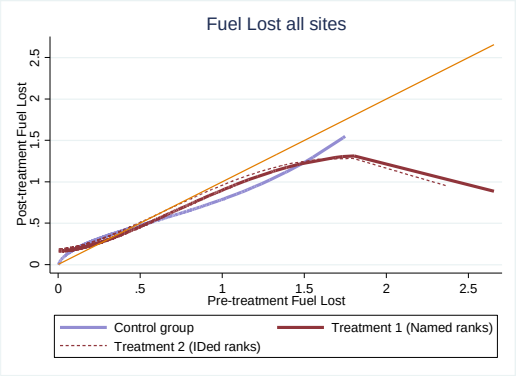
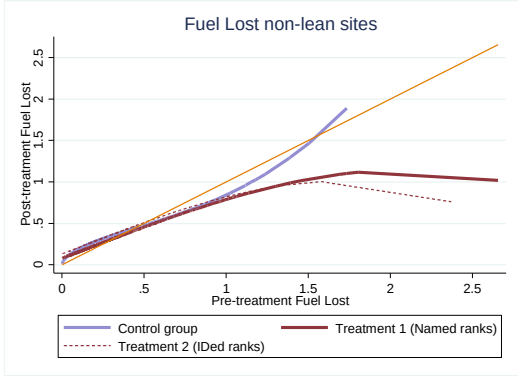
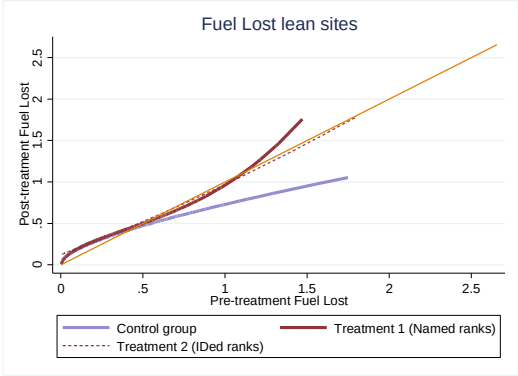


Figure 3d: Fuel Lost







Tables

Table 1: Descriptive statistics

	All Sites				Sample				Control	Treatment 1 (names)	Diff	Treatment 2 (IDs)	Diff
	Mean	Med	Min	Max	Mean	Med	Min	Max	Mean	Mean	p-value	Mean	p-value
Site characteristics													
# sites	275	n/a	n/a	n/a	142	n/a	n/a	n/a	47	50	n/a	45	n/a
Lean status	0.37	n/a	0	1	0.34	n/a	0	1	0.30	0.26	0.681	0.47	0.098
Tractors / site	30.96	22	5	151	24.71	21	5	87	25	25.32	0.924	23.73	0.664
Distance / trip	127.64	127.42	43.76	196.46	128.65	128.5	56.08	200.65	124.08	130.63	0.309	131.24	0.247
Pre-ranking driver performance (site mean)													
Miles per gallon	6.81	6.83	5.35	8.23	6.82	6.80	5.66	8.37	6.76	6.88	0.247	6.82	0.558
Gap score	2.19	2.1	0.57	6.99	2.10	1.94	0.78	6.43	2.18	2.14	0.787	1.98	0.310
Shift score	90.82	91.2	74.78	97.28	91.07	91.53	74.43	97.41	90.77	90.69	0.902	91.79	0.149
Excess idle time	0.12	0.1	0	0.72	0.13	0.11	0.02	0.72	0.12	0.12	0.838	0.14	0.429
Fuel lost	0.33	0.32	0.12	0.74	0.33	0.33	0.14	0.72	0.34	0.35	0.722	0.31	0.185

Table 2: Effect of rankings on all sites

Dependent variable:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment group 1	0.0061	0.0107	-0.6105	-0.6979*	0.0004	0.0007	-0.0007	0.0029
	(0.0213)	(0.0206)	(0.3902)	(0.3736)	(0.0076)	(0.0072)	(0.0077)	(0.0073)
Post*Treatment group 2	0.0429	0.0490**	-0.3728	-0.4428	0.0138*	0.0124*	0.0118	0.0128*
	(0.0262)	(0.0221)	(0.3786)	(0.3622)	(0.0077)	(0.0074)	(0.0082)	(0.0076)
Post	0.0139	0.0061	0.1084	0.2289	0.0216***	0.0188***	0.0079	0.0026
	(0.0168)	(0.0149)	(0.3121)	(0.3016)	(0.0047)	(0.0046)	(0.0054)	(0.0051)
Treatment group 1 (names)	-0.0030	-0.0096	0.1024	0.0324	-0.0048	-0.0116	0.0081	-0.0020
	(0.0445)	(0.0262)	(0.4995)	(0.3689)	(0.0088)	(0.0077)	(0.0118)	(0.0099)
Treatment group 2 (IDs)	-0.0845*	-0.0147	0.8754*	0.0980	0.0008	0.0110	-0.0161	-0.0043
	(0.0477)	(0.0363)	(0.4861)	(0.4372)	(0.0085)	(0.0072)	(0.0129)	(0.0142)
Constant	0.9366***	1.7845***	92.5090***	90.4229***	0.1001***	0.2794***	0.2524***	-0.0842*
	(0.0327)	(0.1247)	(0.3352)	(1.4176)	(0.0056)	(0.0366)	(0.0080)	(0.0476)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	93913	93913	93913	93913	93913	93913	93913	93913
Adjusted R-squared	0.003	0.128	0.003	0.044	0.005	0.050	0.002	0.102

Table 3: Effect of rankings on lean and non-lean sites [suppress pairwise interactions in later tables?]

Dependent variable:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment group 1*Lean	0.1358***	0.1338***	-1.8625**	-1.9657***	0.0394***	0.0352***	0.0593***	0.0538***
	(0.0384)	(0.0383)	(0.7702)	(0.7120)	(0.0136)	(0.0131)	(0.0144)	(0.0127)
Post*Treatment group 2*Lean	0.0251	0.0329	-0.4615	-0.6557	-0.0080	-0.0073	0.0093	0.0131
	(0.0522)	(0.0462)	(0.7701)	(0.7176)	(0.0159)	(0.0152)	(0.0168)	(0.0161)
Post*Treatment group 1	-0.0423*	-0.0375	0.0904	0.0366	-0.0130	-0.0113	-0.0217***	-0.0164**
	(0.0251)	(0.0232)	(0.4018)	(0.3800)	(0.0084)	(0.0080)	(0.0077)	(0.0080)
Post*Treatment group 2	0.0329	0.0333	-0.2777	-0.2295	0.0178	0.0160	0.0079	0.0064
	(0.0392)	(0.0336)	(0.3835)	(0.3777)	(0.0134)	(0.0128)	(0.0116)	(0.0117)
Post*Lean	-0.0303	-0.0269	0.9783	1.0118*	0.0034	0.0027	-0.0098	-0.0090
	(0.0326)	(0.0295)	(0.6547)	(0.6043)	(0.0091)	(0.0090)	(0.0111)	(0.0100)
Treatment group 1*Lean	0.1434*	0.0415	0.5095	0.7454	0.0066	-0.0166	0.0149	0.0066
	(0.0864)	(0.0658)	(0.9945)	(0.7343)	(0.0180)	(0.0163)	(0.0242)	(0.0241)
Treatment group 2*Lean	0.0549	0.0814	-0.7479	-0.3329	-0.0002	0.0141	-0.0022	0.0294
	(0.0895)	(0.0682)	(0.9417)	(0.7676)	(0.0165)	(0.0125)	(0.0250)	(0.0266)
Post	0.0263	0.0180	-0.2856	-0.1838	0.0203***	0.0178***	0.0118*	0.0066
	(0.0222)	(0.0186)	(0.2532)	(0.2425)	(0.0064)	(0.0063)	(0.0065)	(0.0065)
Lean	-0.0937	-0.0814*	0.1299	0.2731	-0.0014	-0.0123	-0.0139	-0.0264
	(0.0590)	(0.0445)	(0.6939)	(0.5931)	(0.0106)	(0.0093)	(0.0160)	(0.0169)
Treatment group 1 (names)	-0.0578	-0.0296	-0.0684	-0.1717	-0.0072	-0.0071	0.0022	-0.0062
	(0.0575)	(0.0337)	(0.6552)	(0.4627)	(0.0116)	(0.0102)	(0.0146)	(0.0131)
Treatment group 2 (IDs)	-0.0996	-0.0503	1.2832**	0.2648	0.0012	0.0049	-0.0124	-0.0174
	(0.0659)	(0.0426)	(0.5405)	(0.5338)	(0.0121)	(0.0104)	(0.0162)	(0.0166)
Constant	0.9741***	1.7811***	92.4570***	90.7710***	0.1007***	0.2767***	0.2579***	-0.0854*
	(0.0453)	(0.1292)	(0.4212)	(1.4685)	(0.0081)	(0.0371)	(0.0098)	(0.0488)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	93913	93913	93913	93913	93913	93913	93913	93913
Adjusted R-squared	0.009	0.130	0.004	0.045	0.007	0.051	0.004	0.104

Notes: SE clustered by SIC, window (5,30), controls include XYZ, winsorization

Table 4: Analysis using date and driver fixed effects

Dependent variable:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment group 1*Lean	0.1026***	0.1067***	-1.9433***	-1.9459***	0.0279**	0.0254*	0.0501***	0.0441***
	(0.0359)	(0.0355)	(0.6079)	(0.6041)	(0.0140)	(0.0132)	(0.0141)	(0.0124)
Post*Treatment group 2*Lean	0.0544	0.0561	-1.2449**	-1.2635**	-0.0087	-0.0069	0.0169	0.0192
	(0.0396)	(0.0389)	(0.5879)	(0.5845)	(0.0157)	(0.0152)	(0.0143)	(0.0136)
Post*Treatment group 1	-0.0424*	-0.0452**	0.3998	0.3974	-0.0133	-0.0109	-0.0231***	-0.0180**
	(0.0229)	(0.0221)	(0.3492)	(0.3484)	(0.0087)	(0.0084)	(0.0079)	(0.0077)
Post*Treatment group 2	0.0112	0.0115	0.0615	0.0674	0.0143	0.0131	0.0042	0.0021
	(0.0304)	(0.0292)	(0.3261)	(0.3252)	(0.0127)	(0.0126)	(0.0099)	(0.0104)
Post*Lean	-0.0368	-0.0415	1.2533**	1.2798***	0.0062	0.0052	-0.0130	-0.0126
	(0.0272)	(0.0271)	(0.4947)	(0.4896)	(0.0094)	(0.0090)	(0.0105)	(0.0092)
Treatment group 1*Lean								
Treatment group 2*Lean								
Post	-0.0049	-0.0027	-0.8095***	-0.8222***	-0.0069	-0.0064	-0.0066	-0.0067
	(0.0216)	(0.0208)	(0.2753)	(0.2741)	(0.0081)	(0.0080)	(0.0082)	(0.0080)
Lean								
Treatment group 1 (names)								
Treatment group 2 (IDs)								
Constant	0.9495***	1.8575***	92.9190***	88.6765***	0.1260***	0.1821***	0.2564***	-0.0576***
	(0.0119)	(0.0484)	(0.1367)	(0.5761)	(0.0057)	(0.0131)	(0.0052)	(0.0195)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	93913	93913	93913	93913	93913	93913	93913	93913
Adjusted R-squared	0.546	0.559	0.601	0.602	0.285	0.298	0.459	0.509

Table 5: Comparison of lean and non-lean sites

	Full sample			Matched sample		
	Non-lean Mean	Lean Mean	Diff p-value	Non-lean Mean	Lean Mean	Diff *p*-value
# sites	94	48	n/a	39	39	n/a
Tractors / site	20.35	33.25	0.000	25.95	27.51	0.581
Distance / trip	128.04	127.53	0.609	128.04	127.53	0.937
Miles per gallon	6.90	6.72	0.039	6.76	6.71	0.602
Gap score	2.14	2.04	0.537	2.00	2.03	0.838
Shift score	90.35	91.55	0.076	91.62	91.66	0.950
Excess idle time	0.12	0.13	0.781	0.12	0.13	0.815
Fuel lost	0.34	0.33	0.473	0.32	0.33	0.753
Eastern region	0.27	0.39	0.155	0.37	0.38	0.865
Central region	0.41	0.37	0.626	0.44	0.38	0.626
Western region	0.32	0.24	0.357	0.20	0.23	0.701
Control group	0.35	0.29	0.480	0.39	0.27	0.245
Treatment group 1	0.39	0.27	0.149	0.32	0.24	0.467
Treatment group 2	0.26	0.44	0.027	0.29	0.49	0.072

Table 7: Matched analysis

Dependent variable:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment group 1*Lean	0.1485***	0.1510***	0.3944	0.0457	0.0523***	0.0522***	0.0585***	0.0582***
	(0.0493)	(0.0433)	(0.7437)	(0.7280)	(0.0172)	(0.0157)	(0.0174)	(0.0153)
Post*Treatment group 2*Lean	0.0034	0.0254	0.7460	0.4590	-0.0203	-0.0151	0.0067	0.0084
	(0.0626)	(0.0532)	(0.6049)	(0.5673)	(0.0192)	(0.0181)	(0.0196)	(0.0193)
Post*Treatment group 1	-0.0623*	-0.0496	-0.7339	-0.6517	-0.0214*	-0.0213*	-0.0238**	-0.0197*
	(0.0343)	(0.0327)	(0.5263)	(0.5547)	(0.0123)	(0.0115)	(0.0110)	(0.0112)
Post*Treatment group 2	0.0350	0.0387	-0.3744	-0.3583	0.0298*	0.0258	0.0073	0.0094
	(0.0503)	(0.0443)	(0.3742)	(0.4068)	(0.0173)	(0.0166)	(0.0149)	(0.0153)
Post*Lean	-0.0435	-0.0425	-0.3965	-0.2911	-0.0017	-0.0050	-0.0139	-0.0126
	(0.0408)	(0.0316)	(0.4778)	(0.4521)	(0.0113)	(0.0106)	(0.0126)	(0.0122)
Treatment group 1*Lean	0.0582	0.0384	0.3169	-0.8576	-0.0011	0.0015	-0.0131	0.0155
	(0.1096)	(0.0934)	(1.0673)	(0.9340)	(0.0253)	(0.0254)	(0.0268)	(0.0328)
Treatment group 2*Lean	-0.0848	0.0842	-0.2996	-1.4198	-0.0109	0.0208	-0.0400	0.0379
	(0.1134)	(0.0825)	(1.0487)	(0.8966)	(0.0202)	(0.0173)	(0.0295)	(0.0306)
Post	0.0496*	0.0347	-0.1958	-0.0739	0.0245**	0.0216**	0.0173**	0.0113
	(0.0296)	(0.0250)	(0.2566)	(0.3057)	(0.0095)	(0.0094)	(0.0082)	(0.0086)
Lean	0.0135	-0.0331	-0.3308	0.7914	0.0022	-0.0156	0.0198	-0.0179
	(0.0839)	(0.0649)	(0.7787)	(0.7736)	(0.0138)	(0.0139)	(0.0189)	(0.0246)
Treatment group 1 (names)	-0.0242	0.0064	0.0517	0.3504	-0.0003	-0.0074	0.0022	0.0078
	(0.0712)	(0.0628)	(0.7174)	(0.6182)	(0.0174)	(0.0192)	(0.0177)	(0.0240)
Treatment group 2 (IDs)	0.0396	-0.0702	0.7376	0.8761	0.0094	0.0044	0.0105	-0.0263
	(0.0813)	(0.0680)	(0.6374)	(0.8571)	(0.0150)	(0.0142)	(0.0207)	(0.0261)
Constant	0.8846***	1.7351***	93.3009***	89.4941***	0.0973***	0.2914***	0.2427***	-0.0876
	(0.0574)	(0.1485)	(0.4419)	(1.6906)	(0.0100)	(0.0356)	(0.0126)	(0.0538)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	60002	60002	60002	60002	60002	60002	60002	60002
Adjusted R-squared	0.007	0.127	0.003	0.050	0.009	0.056	0.004	0.102

Table 8: Matched analysis with date and driver fixed effects

Dependent variable:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment group 1*Lean	0.1088***	0.1092***	-0.1514	-0.1495	0.0360**	0.0350**	0.0453***	0.0435***
	(0.0392)	(0.0380)	(0.6861)	(0.6841)	(0.0172)	(0.0161)	(0.0152)	(0.0135)
Post*Treatment group 2*Lean	0.0118	0.0214	-0.0348	-0.0812	-0.0267	-0.0234	0.0031	0.0043
	(0.0436)	(0.0426)	(0.5283)	(0.5263)	(0.0179)	(0.0175)	(0.0163)	(0.0158)
Post*Treatment group 1	-0.0525*	-0.0514*	-0.4054	-0.4153	-0.0215*	-0.0196*	-0.0231**	-0.0204*
	(0.0308)	(0.0303)	(0.4953)	(0.4945)	(0.0114)	(0.0110)	(0.0114)	(0.0107)
Post*Treatment group 2	0.0317	0.0274	-0.1981	-0.1755	0.0280*	0.0259*	0.0096	0.0080
	(0.0397)	(0.0388)	(0.3252)	(0.3275)	(0.0152)	(0.0152)	(0.0137)	(0.0141)
Post*Lean	-0.0299	-0.0355	0.0595	0.0869	0.0026	0.0006	-0.0078	-0.0086
	(0.0279)	(0.0268)	(0.4627)	(0.4589)	(0.0112)	(0.0109)	(0.0110)	(0.0102)
Treatment group 1*Lean								
Treatment group 2*Lean								
Post	0.0065	0.0106	-0.2165	-0.2353	-0.0070	-0.0059	-0.0018	-0.0019
	(0.0269)	(0.0259)	(0.3157)	(0.3164)	(0.0113)	(0.0110)	(0.0105)	(0.0102)
Lean								
Treatment group 1 (names)								
Treatment group 2 (IDs)								
Constant	0.9295***	1.8671***	93.4125***	89.6393***	0.1297***	0.2003***	0.2558***	-0.0455*
	(0.0161)	(0.0640)	(0.1689)	(0.6491)	(0.0073)	(0.0155)	(0.0072)	(0.0250)
Controls	N	Y	N	Y	N	Y	N	Y
Date and Driver Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
Observations	60002	60002	60002	60002	60002	60002	60002	60002
Adjusted R-squared	0.559	0.572	0.620	0.621	0.289	0.303	0.470	0.521

Table 9: Asymmetric analysis

Table 9a: Triple interaction

Dependent variable: Post-treatment performance	Log(Gap score)	Shift score	Log(Idle Time)	Log(Fuel Lost)
	(1)	(2)	(3)	(4)
Pre-treat Perf*Treatment Group 1*Lean	0.1038	0.2188**	0.3100**	0.3228***
	(0.0874)	(0.1063)	(0.1366)	(0.0971)
Pre-treat Perf*Treatment Group 2*Lean	0.0943	0.1694**	0.0717	0.2830**
	(0.0874)	(0.0848)	(0.1766)	(0.1151)
Pre-treat Perf*Treatment Group 1	-0.0444	-0.0674	-0.1627*	-0.1449***
	(0.0466)	(0.0549)	(0.0888)	(0.0517)
Pre-treat Perf*Treatment Group 2	-0.0167	-0.0465	0.0126	-0.1488**
	(0.0574)	(0.0661)	(0.1116)	(0.0665)
Lean	-0.0088	0.0279	0.0042	0.0017
	(0.0139)	(0.2185)	(0.0038)	(0.0096)
Pre-treat Perf	0.6867***	0.7489***	0.7321***	0.7911***
	(0.0297)	(0.0342)	(0.0674)	(0.0302)
Treatment Group 1	-0.0238**	0.2018	-0.0009	-0.0107
	(0.0111)	(0.1621)	(0.0021)	(0.0070)
Treatment Group 2	-0.0134	0.1203	0.0028	-0.0034
	(0.0127)	(0.1877)	(0.0042)	(0.0069)
Pre-treat Perf*Lean	0.0030	-0.0551	-0.0401	-0.1083
	(0.0605)	(0.0590)	(0.0889)	(0.0771)
Treatment Group 1*Lean	0.0161	-0.0058	-0.0008	0.0033
	(0.0212)	(0.3366)	(0.0054)	(0.0121)
Treatment Group 2*Lean	0.0257	-0.2135	-0.0069	0.0131
	(0.0192)	(0.3015)	(0.0062)	(0.0115)
Constant	0.0189**	0.0885	0.0066***	-0.0119**
	(0.0078)	(0.1251)	(0.0016)	(0.0047)
Observations	2433	2433	2433	2433
Adjusted R-squared	0.574	0.607	0.459	0.551

Table 9b: Split by control and treatments

	Log(Gap score)		Shift score		Log(Idle Time)		Log(Fuel Lost)	
	Control	Rank groups	Control	Rank groups	Control	Rank groups	Control	Rank groups
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Pre-treat Perf*Lean	0.0030	0.1042**	-0.0551	0.1382***	-0.0401	0.1343	-0.1083	0.1923***
	(0.0609)	(0.0437)	(0.0594)	(0.0471)	(0.0895)	(0.1045)	(0.0776)	(0.0532)
Lean	-0.0088	0.0160	0.0279	-0.1254	0.0042	0.0007	0.0017	0.0130**
	(0.0140)	(0.0098)	(0.2199)	(0.1580)	(0.0038)	(0.0029)	(0.0097)	(0.0052)
Pre-treat Perf	0.6867***	0.6540***	0.7489***	0.6896***	0.7321***	0.6652***	0.7911***	0.6440***
	(0.0299)	(0.0300)	(0.0344)	(0.0344)	(0.0679)	(0.0626)	(0.0304)	(0.0345)
Constant	0.0189**	-0.0007	0.0885	0.2562***	0.0066***	0.0069***	-0.0119**	-0.0197***
	(0.0079)	(0.0063)	(0.1259)	(0.0838)	(0.0016)	(0.0018)	(0.0047)	(0.0039)
Observations	898	1535	898	1535	898	1535	898	1535
Adjusted R-squared	0.541	0.592	0.576	0.625	0.430	0.472	0.527	0.564

Appendix Figures and Tables

Figure A1: Asymmetric Treatment Response, site-demeaned

Figure A1a: Gap Score

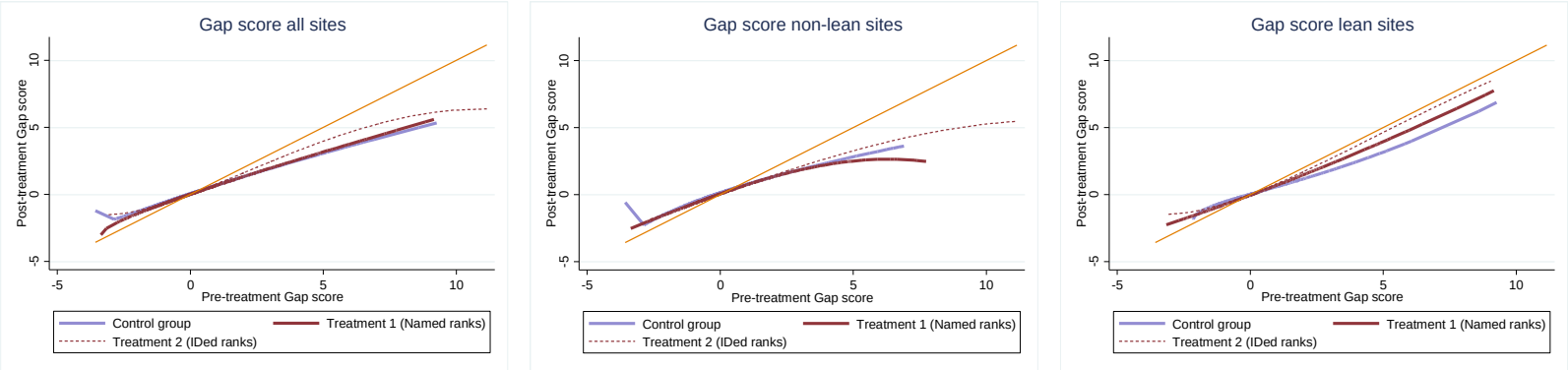


Figure A1b: Shift Score

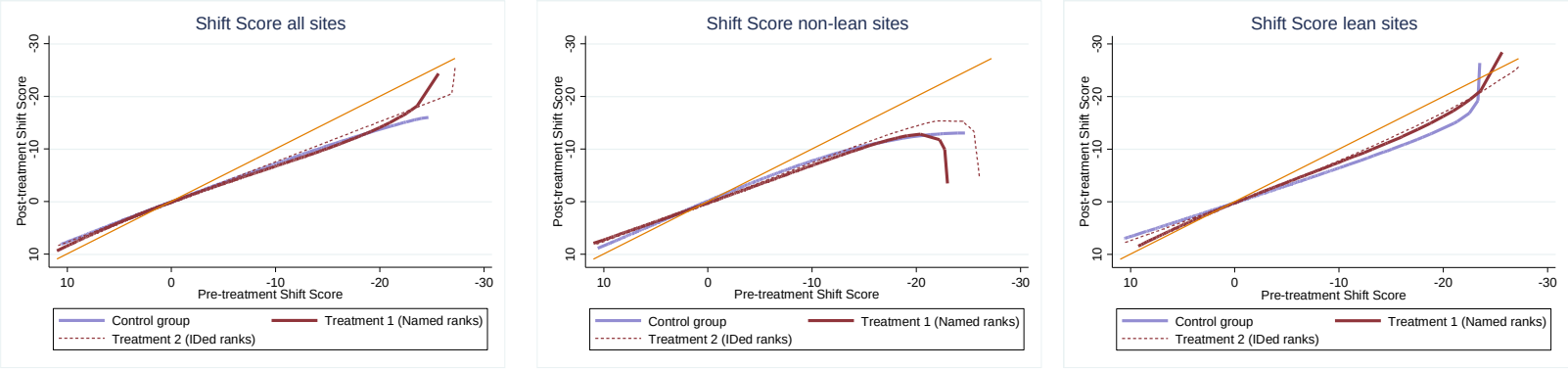


Figure A1c: Idle Time

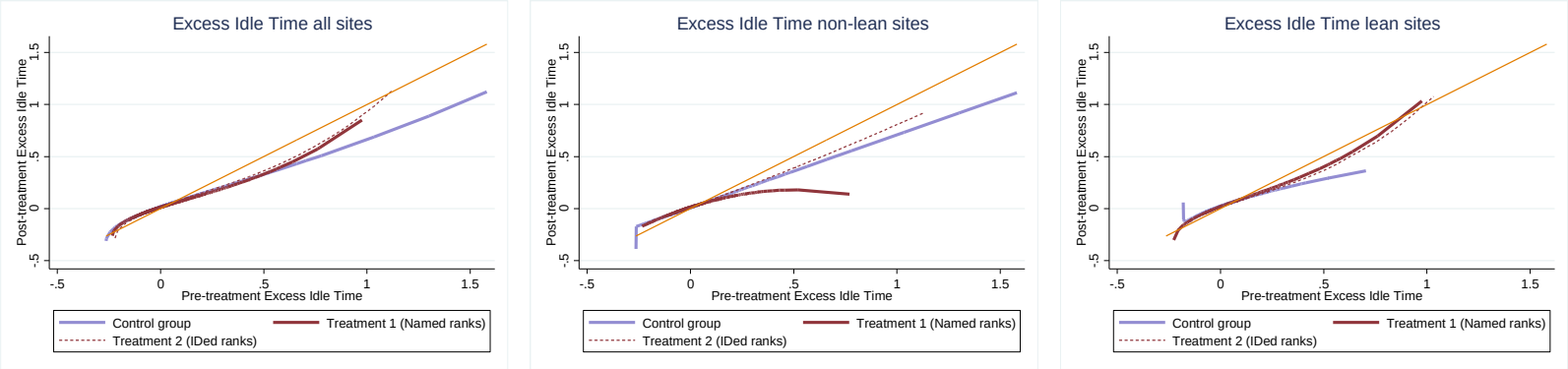


Figure A1d: Fuel Lost

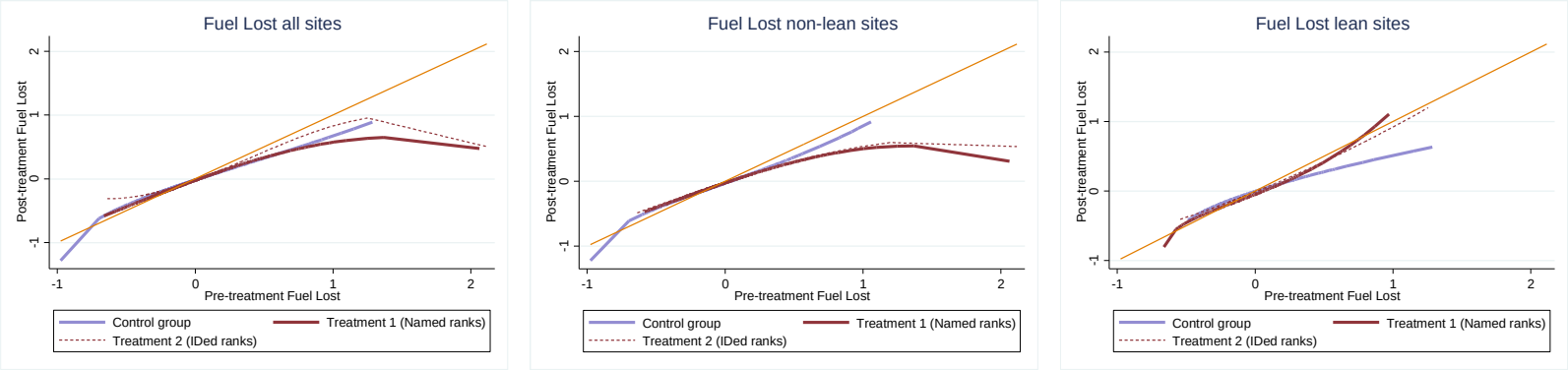


Table A1: Sample construction

Sample construction	Driver-days	Sites
Total driver-days	1,137,192	275
- less early lean sites	(173,461)	(25)
- less late Q3/Q4 2013 lean sites	(130,679)	(36)
- less pre-11/25 rank posting dates	(416,593)	(72)
- less line haul routes	(76,989)	0
- less uncalibrated data	(8,781)	0
Sample	330,689	142
Sample within 5-30 window	93,913	142

Table A2: Descriptive statistics for matched sample

	Matched sample				Control	Treatment 1 (names)	Diff	Treatment 2 (IDs)	Diff
	Mean	Median	Min	Max	Mean	Mean	p-value	Mean	p-value
Site characteristics									
# sites	82	n/a	n/a	n/a	27	23	n/a	32	n/a
Lean status	0.50	n/a	0	1	0.41	0.43	0.849	0.63	0.099
Tractors / site	26.73	22.50	8	61	24.96	29.61	0.213	26.16	0.694
Distance / trip	127.79	126.34	62.08	198.69	128.26	127.66	0.945	127.47	0.918
Pre-ranking driver performance									
Miles per gallon	6.74	6.71	5.67	7.77	6.60	6.80	0.125	6.80	0.083
Gap score	2.01	1.96	0.83	5.26	2.03	2.02	0.993	1.99	0.885
Shift score	91.64	92.25	82.67	97.37	90.92	91.83	0.290	92.10	0.150
Excess idle time	0.12	0.11	0.04	0.68	0.11	0.12	0.800	0.14	0.364
Fuel lost	0.32	0.31	0.15	0.71	0.33	0.31	0.762	0.32	0.554

Table A3: Placebo test

	Actual MPG		Potential MPG	
	(1)	(2)	(3)	(4)
Post*Treatment group 1*Lean	-0.0347***	-0.0223**	0.0111	-0.0560
	(0.0102)	(0.0100)	(0.0602)	(0.0527)
Post*Treatment group 2*Lean	-0.0081	-0.0163	0.0887	0.0679
	(0.0120)	(0.0107)	(0.0595)	(0.0541)
Post*Treatment group 1	0.0065	0.0082	0.0284	0.0589*
	(0.0063)	(0.0062)	(0.0403)	(0.0339)
Post*Treatment group 2	-0.0107	-0.0024	-0.0544	-0.0353
	(0.0091)	(0.0082)	(0.0446)	(0.0348)
Post*Lean	0.0060	0.0081	-0.1129***	-0.0579
	(0.0079)	(0.0074)	(0.0399)	(0.0383)
Treatment group 1*Lean	-0.0153		-0.1943*	
	(0.0174)		(0.1018)	
Treatment group 2*Lean	-0.0204		-0.2488**	
	(0.0165)		(0.1189)	
Post	-0.0001	0.0056	-0.0779***	0.0289
	(0.0050)	(0.0059)	(0.0273)	(0.0373)
Lean	0.0220*		0.2184**	
	(0.0115)		(0.0928)	
Treatment group 1 (names)	0.0075		0.2161***	
	(0.0089)		(0.0678)	
Treatment group 2 (IDs)	0.0089		0.2222***	
	(0.0109)		(0.0819)	
Constant	-0.2591***	-0.0470***	7.0143***	6.7390***
	(0.0310)	(0.0112)	(0.1990)	(0.0264)
Date and Driver FE	N	Y	N	Date
Observations	93913	93913	93913	93913
Adjusted R-squared	0.976	0.985	0.159	0.577

Table A4: Placebo test on matched sample

	Actual MPG		Potential MPG	
	(1)	(2)	(3)	(4)
Post*Treatment group 1*Lean	-0.0341***	-0.0202**	-0.0159	-0.0272
	(0.0102)	(0.0097)	(0.0819)	(0.0695)
Post*Treatment group 2*Lean	-0.0060	-0.0050	0.1391*	0.1413**
	(0.0132)	(0.0115)	(0.0724)	(0.0608)
Post*Treatment group 1	0.0072	0.0096	0.0224	0.0595
	(0.0083)	(0.0078)	(0.0641)	(0.0499)
Post*Treatment group 2	-0.0110	-0.0081	-0.0996*	-0.0812*
	(0.0112)	(0.0105)	(0.0565)	(0.0420)
Post*Lean	0.0068	0.0035	-0.1061*	-0.0846*
	(0.0072)	(0.0061)	(0.0541)	(0.0464)
Treatment group 1*Lean	-0.0234		-0.2250*	
	(0.0242)		(0.1310)	
Treatment group 2*Lean	-0.0177		-0.2563*	
	(0.0209)		(0.1527)	
Post	-0.0027	0.0042	-0.0890**	0.0498
	(0.0058)	(0.0061)	(0.0394)	(0.0489)
Lean	0.0128		0.2627**	
	(0.0160)		(0.1145)	
Treatment group 1 (names)	0.0048		0.2984***	
	(0.0156)		(0.0890)	
Treatment group 2 (IDs)	0.0131		0.2947**	
	(0.0155)		(0.1245)	
Constant	-0.2334***	-0.0518***	6.8689***	6.6435***
	(0.0364)	(0.0136)	(0.2471)	(0.0346)
Date and Driver FE	N	Y	N	Date
Observations	60002	60002	60002	60002
Adjusted R-squared	0.976	0.985	0.159	0.571

Table A5: Instrumental variables analysis

Dependent variable:	log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment group 1*Lean	0.2553**	0.2962**	-4.3968**	-4.7937**	0.0837**	0.0793**	0.1175**	0.1194**
	(0.0997)	(0.1283)	(2.2285)	(2.3630)	(0.0344)	(0.0369)	(0.0502)	(0.0465)
Post*Treatment group 2*Lean	0.0261	0.0325	-0.6834	-0.9888	-0.0281	-0.0260	0.0131	0.0177
	(0.1291)	(0.1184)	(1.6693)	(1.5435)	(0.0459)	(0.0435)	(0.0403)	(0.0392)
Post*Treatment group 1	-0.0755*	-0.0703*	0.1585	0.0624	-0.0229	-0.0206	-0.0382***	-0.0302**
	(0.0441)	(0.0397)	(0.6976)	(0.6735)	(0.0145)	(0.0139)	(0.0138)	(0.0137)
Post*Treatment group 2	0.0859	0.0880	-0.7348	-0.6465	0.0451	0.0391	0.0204	0.0178
	(0.1074)	(0.0973)	(1.0684)	(0.9919)	(0.0433)	(0.0407)	(0.0318)	(0.0317)
Post*Lean	-0.0303	-0.0300	0.9783	1.0158*	0.0034	0.0033	-0.0098	-0.0101
	(0.0325)	(0.0295)	(0.6523)	(0.6034)	(0.0091)	(0.0091)	(0.0111)	(0.0101)
Treatment group 1*Lean	0.3318	0.0873	1.3078	1.9110	0.0111	-0.0518	0.0422	0.0123
	(0.2096)	(0.1675)	(2.3074)	(2.0494)	(0.0421)	(0.0497)	(0.0570)	(0.0609)
Treatment group 2*Lean	0.1667	0.2357	-2.2287	-1.2104	-0.0010	0.0309	0.0028	0.0817
	(0.1887)	(0.1840)	(2.2929)	(1.8750)	(0.0384)	(0.0349)	(0.0540)	(0.0701)
Post	0.0263	0.0191	-0.2856	-0.1774	0.0203***	0.0180***	0.0118*	0.0069
	(0.0221)	(0.0186)	(0.2523)	(0.2430)	(0.0064)	(0.0064)	(0.0065)	(0.0065)
Lean	-0.0937	-0.0900*	0.1299	0.3012	-0.0014	-0.0130	-0.0139	-0.0283
	(0.0587)	(0.0467)	(0.6914)	(0.6381)	(0.0106)	(0.0100)	(0.0159)	(0.0174)
Treatment group 1 (names)	-0.1011	-0.0597	-0.1197	-0.3691	-0.0126	-0.0135	0.0039	-0.0136
	(0.0982)	(0.0674)	(1.1465)	(0.9449)	(0.0200)	(0.0210)	(0.0257)	(0.0262)
Treatment group 2 (IDs)	-0.2553*	-0.1596	3.2883*	0.8652	0.0030	0.0029	-0.0317	-0.0540
	(0.1470)	(0.1035)	(1.7225)	(1.3275)	(0.0312)	(0.0298)	(0.0398)	(0.0387)
Constant	0.9741***	1.7826***	92.4570***	90.8150***	0.1007***	0.2884***	0.2579***	-0.0859*
	(0.0452)	(0.1297)	(0.4197)	(1.4618)	(0.0081)	(0.0371)	(0.0098)	(0.0484)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	93913	93913	93913	93913	93913	93913	93913	93913
Adjusted R-squared	0.011	0.126		0.041	0.002	0.045	0.002	0.101

Table A6: Instrumental variable analysis on matched sample

Dependent variable:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment group 1*Lean	0.2880**	0.3091**	0.5184	-0.1058	0.1055***	0.1065**	0.1208**	0.1179***
	(0.1145)	(0.1201)	(1.6522)	(1.6767)	(0.0406)	(0.0414)	(0.0496)	(0.0427)
Post*Treatment group 2*Lean	-0.1089	-0.0627	2.0143	1.7536	-0.1269	-0.1009	-0.0109	-0.0108
	(0.2861)	(0.2453)	(2.1440)	(2.2801)	(0.1382)	(0.1190)	(0.0778)	(0.0798)
Post*Treatment group 1	-0.1185*	-0.0989*	-1.3615	-1.1202	-0.0399*	-0.0415*	-0.0441**	-0.0391**
	(0.0639)	(0.0553)	(1.2229)	(1.2327)	(0.0216)	(0.0214)	(0.0186)	(0.0188)
Post*Treatment group 2	0.1779	0.1703	-1.4758	-1.5799	0.1425	0.1166	0.0379	0.0406
	(0.2785)	(0.2386)	(2.0199)	(2.1887)	(0.1375)	(0.1185)	(0.0745)	(0.0766)
Post*Lean	-0.0435	-0.0424	-0.3965	-0.2954	-0.0017	-0.0042	-0.0139	-0.0125
	(0.0405)	(0.0323)	(0.4749)	(0.4401)	(0.0112)	(0.0106)	(0.0125)	(0.0127)
Treatment group 1*Lean	0.1259	0.1508	0.8134	-2.2238	-0.0030	0.0117	-0.0306	0.0677
	(0.2426)	(0.2461)	(2.2628)	(2.4720)	(0.0545)	(0.0591)	(0.0582)	(0.0892)
Treatment group 2*Lean	-0.2500	0.4302	-2.3961	-5.5245	-0.0432	0.1004	-0.0973	0.1904
	(0.4089)	(0.3118)	(3.3864)	(3.3759)	(0.0767)	(0.0801)	(0.0986)	(0.1237)
Post	-0.0425	0.0035	0.0908	0.5247	-0.0006	-0.0259	0.0039	0.0126
	(0.1235)	(0.1413)	(1.2531)	(1.4532)	(0.0303)	(0.0413)	(0.0312)	(0.0559)
Lean	0.1700	-0.2977	3.1711	3.4086	0.0405	-0.0462	0.0451	-0.1169
	(0.3858)	(0.1901)	(3.0445)	(2.1764)	(0.0730)	(0.0607)	(0.0923)	(0.0755)
Treatment group 1 (names)	0.0496*	0.0338	-0.1958	-0.0597	0.0245***	0.0214**	0.0173**	0.0109
	(0.0294)	(0.0254)	(0.2550)	(0.3086)	(0.0095)	(0.0094)	(0.0082)	(0.0089)
Treatment group 2 (IDs)	0.0135	-0.0663	-0.3308	1.0935	0.0022	-0.0217	0.0198	-0.0346
	(0.0834)	(0.0801)	(0.7739)	(0.9326)	(0.0137)	(0.0167)	(0.0187)	(0.0309)
Constant	0.8846***	1.7607***	93.3009***	89.0525***	0.0973***	0.2958***	0.2427***	-0.0746
	(0.0570)	(0.1562)	(0.4391)	(1.6999)	(0.0100)	(0.0379)	(0.0125)	(0.0553)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	60002	60002	60002	60002	60002	60002	60002	60002
Adjusted R-squared	.	0.117	.	0.046	.	0.048	0.004	0.088

Table A7: Effect of rankings on lean and non-lean sites - long window

Dependent variable:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment group 1*Lean	0.0966**	0.0963**	-1.6616	-1.4921	0.0259	0.0252	0.0465***	0.0373***
	(0.0441)	(0.0429)	(1.0370)	(1.0080)	(0.0163)	(0.0156)	(0.0156)	(0.0140)
Post*Treatment group 2*Lean	0.0263	0.0297	-1.5130	-1.4498	-0.0036	0.0001	0.0130	0.0131
	(0.0486)	(0.0464)	(0.9732)	(0.9436)	(0.0183)	(0.0167)	(0.0185)	(0.0161)
Post*Treatment group 1	-0.0414	-0.0526**	0.6129	0.6155	-0.0116	-0.0114	-0.0259***	-0.0201**
	(0.0270)	(0.0248)	(0.6369)	(0.6179)	(0.0082)	(0.0073)	(0.0093)	(0.0090)
Post*Treatment group 2	0.0289	0.0295	-0.0243	-0.0978	0.0168	0.0141	0.0065	0.0069
	(0.0354)	(0.0325)	(0.7387)	(0.7135)	(0.0127)	(0.0117)	(0.0133)	(0.0113)
Post*Lean	-0.0280	-0.0332	0.3156	0.3168	0.0066	0.0023	-0.0124	-0.0115
	(0.0328)	(0.0317)	(0.7552)	(0.7297)	(0.0121)	(0.0111)	(0.0117)	(0.0103)
Treatment group 1*Lean	0.1407	0.0377	0.3884	0.9014	0.0100	-0.0046	0.0105	0.0076
	(0.0870)	(0.0675)	(1.3413)	(1.1558)	(0.0168)	(0.0161)	(0.0227)	(0.0246)
Treatment group 2*Lean	0.0653	0.0556	-0.1648	0.8462	-0.0000	0.0080	0.0007	0.0194
	(0.0852)	(0.0692)	(1.1586)	(1.0033)	(0.0145)	(0.0157)	(0.0230)	(0.0269)
Post	0.0471**	0.0532***	1.6888***	1.6883***	0.0170***	0.0189***	0.0229***	0.0198***
	(0.0196)	(0.0167)	(0.4700)	(0.4476)	(0.0060)	(0.0053)	(0.0065)	(0.0058)
Lean	-0.0936*	-0.0628	0.7248	0.3364	-0.0034	-0.0166	-0.0121	-0.0200
	(0.0561)	(0.0487)	(0.8861)	(0.7853)	(0.0096)	(0.0107)	(0.0148)	(0.0180)
Treatment group 1 (names)	-0.0499	-0.0072	-0.0830	-0.4096	-0.0081	-0.0076	0.0038	0.0007
	(0.0553)	(0.0325)	(0.9339)	(0.7730)	(0.0101)	(0.0095)	(0.0131)	(0.0124)
Treatment group 2 (IDs)	-0.1021*	-0.0355	1.4287*	0.2472	-0.0011	0.0031	-0.0164	-0.0121
	(0.0614)	(0.0438)	(0.8018)	(0.7815)	(0.0104)	(0.0109)	(0.0142)	(0.0170)
Constant	0.9776***	1.7193***	90.3028***	87.4869***	0.0953***	0.2802***	0.2588***	-0.1257***
	(0.0428)	(0.1134)	(0.5702)	(1.3161)	(0.0071)	(0.0289)	(0.0085)	(0.0440)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	310084	310084	310084	310084	310084	310084	310084	310084
Adjusted R-squared	0.008	0.128	0.010	0.052	0.006	0.048	0.004	0.107

Table A8: Date and driver fixed effects - long window

Dependent variable:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment group 1*Lean	0.0829*	0.0822*	-1.4671	-1.4290	0.0264*	0.0228	0.0375**	0.0322**
	(0.0423)	(0.0426)	(1.0750)	(1.0707)	(0.0154)	(0.0148)	(0.0154)	(0.0144)
Post*Treatment group 2*Lean	0.0350	0.0333	-1.6876*	-1.6902*	-0.0026	-0.0013	0.0127	0.0154
	(0.0399)	(0.0402)	(1.0178)	(1.0144)	(0.0152)	(0.0148)	(0.0147)	(0.0139)
Post*Treatment group 1	-0.0516**	-0.0534**	0.7238	0.7138	-0.0095	-0.0074	-0.0251**	-0.0209**
	(0.0259)	(0.0254)	(0.6759)	(0.6738)	(0.0070)	(0.0069)	(0.0097)	(0.0097)
Post*Treatment group 2	0.0095	0.0111	-0.0248	-0.0191	0.0123	0.0107	0.0037	0.0005
	(0.0281)	(0.0276)	(0.7814)	(0.7782)	(0.0108)	(0.0105)	(0.0098)	(0.0096)
Post*Lean	-0.0310	-0.0319	0.4345	0.4371	0.0020	0.0023	-0.0128	-0.0119
	(0.0283)	(0.0288)	(0.7643)	(0.7586)	(0.0097)	(0.0093)	(0.0104)	(0.0096)
Treatment group 1*Lean								
Treatment group 2*Lean								
Post	-0.0031	-0.0047	1.0660**	1.0748**	0.0069	0.0069	-0.0033	-0.0025
	(0.0194)	(0.0187)	(0.5220)	(0.5212)	(0.0074)	(0.0074)	(0.0065)	(0.0065)
Lean								
Treatment group 1 (names)								
Treatment group 2 (IDs)								
Constant	0.9637***	1.8205***	91.4879***	86.2097***	0.1193***	0.1670***	0.2628***	-0.0866***
	(0.0126)	(0.0411)	(0.3789)	(0.6108)	(0.0057)	(0.0105)	(0.0054)	(0.0184)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	310084	310084	310084	310084	310084	310084	310084	310084
Adjusted R-squared	0.515	0.527	0.526	0.529	0.264	0.277	0.430	0.483

Table A9: Matched analysis with date and driver fixed effects - long window

Dependent variable:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment group 1*Lean	0.0990*	0.1004*	-0.5058	-0.4726	0.0445**	0.0389**	0.0464**	0.0378*
	(0.0549)	(0.0541)	(1.4009)	(1.3983)	(0.0177)	(0.0175)	(0.0199)	(0.0194)
Post*Treatment group 2*Lean	0.0031	0.0099	-1.1112	-1.1595	-0.0131	-0.0117	0.0040	0.0025
	(0.0466)	(0.0460)	(1.4296)	(1.4241)	(0.0156)	(0.0156)	(0.0165)	(0.0166)
Post*Treatment group 1	-0.0566	-0.0582	-0.2796	-0.2884	-0.0203**	-0.0179*	-0.0290**	-0.0249*
	(0.0361)	(0.0354)	(0.7970)	(0.7936)	(0.0093)	(0.0091)	(0.0137)	(0.0135)
Post*Treatment group 2	0.0319	0.0299	-0.2026	-0.1741	0.0198	0.0174	0.0095	0.0071
	(0.0359)	(0.0353)	(1.0579)	(1.0549)	(0.0134)	(0.0134)	(0.0124)	(0.0123)
Post*Lean	-0.0375	-0.0439	0.0202	0.0464	-0.0054	-0.0041	-0.0176	-0.0126
	(0.0354)	(0.0341)	(1.1392)	(1.1336)	(0.0082)	(0.0084)	(0.0128)	(0.0130)
Treatment group 1*Lean								
Treatment group 2*Lean								
Post	0.0041	0.0048	1.5634**	1.5626**	0.0112	0.0108	0.0007	-0.0002
	(0.0245)	(0.0231)	(0.6650)	(0.6673)	(0.0097)	(0.0096)	(0.0086)	(0.0086)
Lean								
Treatment group 1 (names)								
Treatment group 2 (IDs)								
Constant	0.9436***	1.7822***	91.8457***	86.7289***	0.1204***	0.1765***	0.2631***	-0.0862***
	(0.0167)	(0.0558)	(0.5084)	(0.7706)	(0.0074)	(0.0129)	(0.0069)	(0.0233)
Controls	N	Y	N	Y	N	Y	N	Y
Date and Driver FE	Y	Y	Y	Y	Y	Y	Y	Y
Observations	198831	198831	198831	198831	198831	198831	198831	198831
Adjusted R-squared	0.519	0.531	0.527	0.529	0.261	0.274	0.437	0.488

Table A10: Matched analysis - long window

Dependent variable:	Log(Gap Score)		Shift Score		Log(Idle Time)		Log(Fuel Lost)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post*Treatment group 1*Lean	0.1230**	0.1366**	-0.6115	-0.5885	0.0573***	0.0536***	0.0645***	0.0528***
	(0.0570)	(0.0522)	(1.4459)	(1.3966)	(0.0186)	(0.0176)	(0.0196)	(0.0187)
Post*Treatment group 2*Lean	0.0116	0.0290	-0.9394	-0.9843	-0.0025	0.0012	0.0125	0.0097
	(0.0582)	(0.0550)	(1.4392)	(1.3895)	(0.0188)	(0.0172)	(0.0227)	(0.0199)
Post*Treatment group 1	-0.0563	-0.0702**	-0.2320	-0.1899	-0.0238*	-0.0228*	-0.0333***	-0.0281**
	(0.0365)	(0.0324)	(0.7581)	(0.7241)	(0.0129)	(0.0116)	(0.0123)	(0.0118)
Post*Treatment group 2	0.0429	0.0304	-0.1244	-0.1089	0.0216	0.0176	0.0063	0.0070
	(0.0452)	(0.0429)	(1.0250)	(0.9792)	(0.0168)	(0.0156)	(0.0178)	(0.0150)
Post*Lean	-0.0524	-0.0629*	-0.2213	-0.1467	-0.0148	-0.0152	-0.0250	-0.0197
	(0.0412)	(0.0370)	(1.1712)	(1.1342)	(0.0105)	(0.0091)	(0.0151)	(0.0141)
Treatment group 1*Lean	0.0791	0.0207	0.0990	-0.6190	0.0041	0.0015	-0.0115	0.0110
	(0.1113)	(0.0900)	(1.5884)	(1.4769)	(0.0229)	(0.0221)	(0.0247)	(0.0317)
Treatment group 2*Lean	-0.0633	0.0739	0.9075	-0.0142	-0.0068	0.0076	-0.0311	0.0371
	(0.1058)	(0.0798)	(1.5747)	(1.3542)	(0.0176)	(0.0160)	(0.0265)	(0.0301)
Post	0.0626**	0.0735***	1.7184***	1.6373***	0.0265***	0.0261***	0.0308***	0.0266***
	(0.0277)	(0.0229)	(0.6430)	(0.6093)	(0.0087)	(0.0079)	(0.0093)	(0.0083)
Lean	0.0007	-0.0004	-0.3159	0.3830	-0.0024	-0.0073	0.0173	-0.0074
	(0.0786)	(0.0651)	(1.2546)	(1.1134)	(0.0124)	(0.0114)	(0.0168)	(0.0246)
Treatment group 1 (names)	-0.0248	0.0284	0.7930	0.7404	-0.0028	-0.0092	0.0014	0.0145
	(0.0701)	(0.0604)	(1.0287)	(0.9204)	(0.0149)	(0.0165)	(0.0153)	(0.0231)
Treatment group 2 (IDs)	0.0266	-0.0850	0.7728	1.0107	0.0050	-0.0025	0.0024	-0.0339
	(0.0761)	(0.0630)	(1.0768)	(1.1107)	(0.0126)	(0.0126)	(0.0181)	(0.0240)
Constant	0.8986***	1.5815***	91.0687***	88.5450***	0.0929***	0.2807***	0.2459***	-0.1567***
	(0.0562)	(0.1392)	(0.7216)	(1.5272)	(0.0086)	(0.0313)	(0.0106)	(0.0503)
Controls	N	Y	N	Y	N	Y	N	Y
Observations	198831	198831	198831	198831	198831	198831	198831	198831
Adjusted R-squared	0.008	0.130	0.006	0.051	0.008	0.051	0.005	0.108