UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

An Analysis Framework for Evaluation of Traffic Compliance Measures

Permalink

<https://escholarship.org/uc/item/4pb066qd>

Author

Campbell, Robert Jay

Publication Date

2012

Peer reviewed|Thesis/dissertation

An Analysis Framework for Evaluation of Traffic Compliance Measures

by

Robert Jay Campbell

A dissertation submitted in partial satisfaction of the
requirements for the degree of

Doctor of Philosophy

in

Engineering - Civil and Environmental Engineering

in the

Graduate Division

of the

University of California at Berkeley

Committee in charge:

Professor Alex Skabardonis, Chair

Professor Robert Leachman

Professor Samer Madanat

Fall 2012

ABSTRACT

An Analysis Framework for Evaluation of Traffic Compliance Measures

by

Robert Jay Campbell

Doctor of Philosophy in Engineering - Civil and Environmental Engineering

University of California, Berkeley

Professor Alex Skabardonis, Chair

Agencies and practitioners are often testing new and innovative strategies for improving driver compliance with traffic regulations. However, in evaluating these strategies, researchers often rely on simple before-and-after methods that suffer from several flaws that can result in misleading results and an inaccurate assessment of a strategy's effectiveness. Specifically, such studies frequently omit control groups to account for other factors that may influence driver behavior aside from the experimental change. Furthermore, these studies often focus on only one compliance measure, and their results are often poorly suited for making comparisons to other compliance strategies that have been evaluated through other before-and-after studies due to the unique details of the experimental sites chosen in each case. Finally, analyses based on the traditional before-and-after approach do not properly account for the period of instability following the experimental change, nor do they make any attempt to characterize it—rather, these studies typically rely on assumptions about how long the instability will last beforehand (and subsequently ignore this period), or fail to account for it at all.

In this dissertation, we examine these flaws and propose a framework that avoids or corrects for them. Among the key features of our proposed framework are a model to describe the driver response to an experimental change (e.g., the increase in compliance following the implementation of a compliance strategy), the inclusion of a baseline prediction model that incorporates control group compliance rates along with other relevant covariates to project what the behavior of the experimental group would have been in the absence of the experimental change, and a measure of effectiveness based on the estimated long-term performance of the compliance strategy after accounting for the period of instability immediately following the experimental change. The framework incorporates the previously documented Novelty Effect, which refers to a short-term boost in compliance due to the novelty of the change, and combines it with a Driver Awareness or driver learning effect, which describes the tendency of the behavioral response (e.g., the boost in compliance) to occur gradually after the experimental change—rather than instantaneously after it—as a result of users taking some time to become aware of the change and to respond appropriately to it. The result is a characteristic driver response curve that is initially increasing after the experimental change, rather than decreasing as is conventionally assumed. When we take a detailed look at compliance data following the implementation of two compliance strategies, we find that the data support this pattern of initially

increasing compliance predicted by our framework.

To illustrate the use of this framework, we consider the case of drivers failing to yield on a series of freeway entrance ramps along Interstate 10 in Los Angeles. The problem occurs at several locations along the freeway, with each site having the same general ramp design but a different set of users. Consequently, the entrance ramps are well suited for the application of our framework, as different compliance strategies can be implemented at various locations and comparisons made regarding the resultant improvements in compliance in each case. Furthermore, a subset of the locations can be kept unchanged, to function as control groups for our analyses.

The first compliance strategy to be analyzed was increasing the size of the signage used. Specifically, the size of the Yield sign was increased to from 36 inches across to 48 inches, which increased its viewable area by 78%. After identifying the best prediction model for the behavior of drivers at this experimental location using the pre-change (or “before”) data, the baseline compliance predictions were made for the post-change weeks. By comparing the observed levels of compliance to these predicted rates, we could then evaluate the improvement in compliance week by week in the “after” period due to the larger sign. Through curve-fitting techniques, we identified the driver response curve that best characterized the observed degree of improvement by week after the change, and from this we concluded that the larger Yield sign resulted in a 22% reduction in non-compliance at the on-ramp, with a 95% confidence interval of 10% to 26%. This measure of effectiveness excludes an estimated 11% of drivers who are deliberately noncompliant at this location and would not be expected to respond to anything apart from increased enforcement. The Driver Awareness effect was found to be significant in the model for the larger Yield sign at a 5% level, although the Novelty Effect was not.

The second experimental change we considered was the widening of the shoulder on one of the entrance ramps, which was found to have a slightly negative impact on driver compliance as it gave some drivers additional room to negotiate inappropriate merges into the stream of cross-traffic. Using our framework, we found that a wider shoulder resulted in a 16% decrease in compliance. In this case, both the Novelty Effect and Driver Awareness effect were significant on a 5% level.

Finally, we examined the effect of a Yield Line on driver compliance and found it to be 27% effective, with a 95% confidence interval of between 17% and 34%. This measure of effectiveness excludes an estimated 14% of drivers who were deliberately non-compliant at this location. Both the Driver Awareness effect and the Novelty Effect were significant on a 5% level in the Yield Line driver response model.

From our results for the larger Yield sign and the Yield Line, we conclude that the Yield Line was 23% more effective than the larger signage at reducing non-compliance at the yield-controlled entrance ramps in the long term, although these results were obtained for a specific set of environmental conditions: midday Sunday traffic in an uncongested suburban environment, excluding rainy days.

In our analysis of the Larger Sign data and the Yield Line data, we found that the durations of the Novelty Effect and the Driver Awareness effect were best described by Normal distributions, although the Driver Awareness effect was also reasonably modeled by a Uniform distribution. Thus, as part of a sensitivity analysis, we explore the effect of using a Uniform distribution in place of a Normal distribution to describe the duration of the Driver Awareness effect.

We finish our analysis by comparing of our measures of effectiveness to the estimates we would have obtained if we had used a more conventional before-and-after approach instead of our framework, to emphasize the improvements achieved by the latter. We also explore alternate versions of the baseline prediction model to confirm that the form we have chosen is optimal with respect to accurate predictions of compliance rates at the experimental locations.

Our results show that the traditionally-held notion of a driver response (following the implementation of a compliance strategy) that starts off strong and tapers off over time is inaccurate, and that the true driver response includes an initial period of increasing compliance instead. This has important consequences for analyses of compliance strategies, as it indicates that data immediately following a strategy's implementation is likely to under-represent, rather than over-represent, its long term effectiveness.

As more studies are carried out following the framework described here, agencies and practitioners will have a growing library of data to reference regarding the durations of the Novelty Effect and Driver Awareness effect for various kinds of compliance strategies. Such information could enable researchers to establish guidelines regarding appropriate “waiting periods” that must be observed before the driver response may be expected to stabilize. With information regarding the time to stability for a particular experimental change, agencies would be able to perform more simplified analyses of long-term effectiveness without needing to fit driver response curves to the data—in these cases, they would simply exclude the unstable portion of the “after” period instead. However, even in these situations, baseline prediction models would still be needed to ensure that other external factors affecting driver behavior are properly controlled for.

Although we have applied our proposed framework to evaluate compliance strategies on a set of yield-controlled ramp merges, it may also be applied to a wide range of driver compliance situations, such as yielding at pedestrian crossings, speed limit obedience, and proper anticipation of roadway features such as speed humps or intersections. Additional research will be needed to reveal the suitability of our framework (and our characteristic driver response curves) for describing the driver response to strategies implemented in these situations and others. Furthermore, additional research may also provide insight into the extent to which our findings regarding Yield Lines and larger signage can be generalized to other contexts, such as inclement weather or night-time driving.

Dedicated to my wonderful family:

Grandma and Grandpa Campbell
Grandma and Grandpa Ito
Kevin

and most of all,
my Mom and Dad

TABLE OF CONTENTS

1. INTRODUCTION	1
Outline.	2
2. LITERATURE REVIEW	2
3. METHODOLOGY	4
Site selection.	4
Creating a Driver Behavior Prediction Model	5
Comparing Predicted and Observed Driver Response	6
Modeling the Driver Response over Time	6
Novelty Persistence Curve	10
Measure of Effectiveness (MOE)	10
Curve-Fitting Parameters	11
Error Term.	11
Adjustments for Compliance Strategies that do not address enforcement issues.	12
4. DATA COLLECTION AND PROCESSING	13
Background.	13
Data Source.	15
Data Overview	15
Data Collection Equipment.	19
Defining a Proper Yield.	21
Camera Angles and Details	22
Theoretical and Practical Considerations	26
Considerations for Specific Cases.	28
Data Analysis	29
Processing the Video Data	31
5. APPLYING THE ANALYSIS FRAMEWORK	33
Initial Survey.	33
Experimental Changes to the Ramps	33
Control Ramps and Other Covariates	34
Ramp User Pools.	37
Prediction Model Formulation	39
Driver Response Curve	39
Estimating β'_n	41
Analysis of Larger Sign	42
<i>Prediction Model Formulation</i>	42
<i>Estimate for β'_n</i>	43
<i>Results</i>	44
<i>Model Parameters</i>	44
<i>Significance of Parameters</i>	46

Analysis of Ramp Rehabilitation	49
<i>Prediction Model Formulation</i>	49
<i>Alternate Prediction Model</i>	51
<i>Prediction Model Results</i>	52
<i>Modified Driver Response Model</i>	54
<i>Estimate for β'_n</i>	55
<i>Preliminary Results for Driver Response to Rehabilitated Ramp</i>	56
<i>Preliminary Model Parameters</i>	56
<i>Final Results for Driver Response to Rehabilitated Ramp</i>	59
Analysis of Yield Line	59
<i>Preliminary Results</i>	59
<i>Checking our Assumption of Normally Distributed Residuals</i>	62
<i>Final Results</i>	68
<i>Significance of Parameters</i>	71
6. UNIFIED RESULTS, SENSITIVITY ANALYSIS, AND ALTERNATE APPROACHES . .	73
Revisiting the Ramp Rehabilitation and Yield Line Analyses	76
Updated Results for the Ramp Rehabilitation and Yield Line	78
Confidence Intervals for MOE	82
Assuming $\beta_3 = 0.82$ with a Uniform distribution for $f_n(t_n)$	82
Other Prediction Models	90
<i>Disaggregate by Hour</i>	90
<i>Arrival Situation Model</i>	91
<i>Driver Choice Model for Location 8</i>	95
Logit choice probabilities	95
Covariates	96
Optimal Choice Model	96
<i>Driver Choice Model for Location 5</i>	97
Optimal Choice Model	99
Conclusions	100
Comparison to Traditional Before-and-After Approach	100
<i>Larger Signage</i>	100
<i>Yield Line</i>	102
7. DISCUSSION	103
Summary of Study Findings	104
Comparisons to Other Regions	105
Applications	105
Future Work	107
Availability of Raw Data	110
8. REFERENCES	111
9. APPENDICES	113
A. Initial research proposal submitted to Caltrans in November 2008.	113

B. Proposal for implementation of the second and third compliance strategies.	118
C. Sample contract with a homeowner for placement of equipment for a year	120
D. Windows Video Processing Tool for yielding events (instructions)	123
E. Mac OS X Video Processing Tool for yielding events (instructions).	126
F. Announcement for research assistants to assist with video processing.	130
G. Windows Traffic Counts Tool for on-ramp volumes (instructions).	132
H. Data tables for Locations 5 and 8 (for analyses of compliance strategies)	135
I. Details of the Choice Model analysis for Locations 5 and 8.	140

LIST OF FIGURES

Figure	Description	Page
1	Driver Awareness and Novelty Persistence curves (examples)	7
2	Typical freeway-arterial interchange design	14
3	Corridor layout and prevalence of compliance problem	14
4	Summary of video data collected between Oct 2009 and Sept 2010	16
5	Summary of video data collected between Oct 2010 and July 2011	17
6	Summary of video data collected between July 2011 and Nov 2012	18
7	On-ramp volumes at Location 5 before, during, and after the Del Mar Ave detour ...	20
8	Camera stands used to monitor on-ramp traffic from adjacent private property	20
9	Camera angles and thresholds used for yielding criteria	23
10	Video processing tool developed to keep track of yielding events in each clip.	30
11	Superimposed guides for the yellow-path and red-path thresholds at Location 8	32
12	Interpretation of the superimposed guides at Location 5	32
13	Photos of the compliance strategies implemented at Locations 1, 5, and 8	35
14	Demographic information for the area surrounding the study corridor	38
15	Observed and predicted compliance by week at Location 8 (larger sign)	45
16	Percent reduction in violation rate by week at Location 8 (larger sign)	45
17	Driver Response Curves excl. Driver Awareness or Novelty Effects, Location 8	48
18	Condition of Location 6 before and after rehabilitation occurred in 2010.	50
19	Observed and predicted compliance by week at Location 5 (ramp rehabilitation) ...	53
20	Percent reduction in violation rate by week at Location 5 (ramp rehabilitation).	57
21	Updated reduction in violation rate by week at Location 5 (ramp rehabilitation).	60
22	Driver Response Curves excl. Awareness or Novelty Effects (ramp rehabilitation) ...	60
23	Observed and predicted compliance by week at Location 5 (Yield Line)	61
24	Percent reduction in violation rate by week at Location 5 (Yield Line)	63
25	Q-Q Plots comparing Driver Response Curve residuals to a Normal distribution ...	67
26	Percent reduction in violation rate by week (Yield Line) with adjusted error terms ...	72
27	Driver Response Curves excl. Driver Awareness or Novelty Effects (Yield Line) ...	74
28	Percent reduction in violation rate by week (Larger Sign) using Unified solution ...	77
29	Pct. reduction in violations by week (ramp rehabilitation) using Unified solution ...	79
30	Percent reduction in violation rate by week (Yield Line) based on Unified solution ..	79
31	Pct. reduction in violations by week (Larger Sign) using alternate Unified solution ..	87
32	Comparing the Hourly prediction model to the (Risk-Minimized) Aggregate model ..	92
33	Comparing the Arrival Situation prediction model to the Aggregate model	94
34	Comparing the Choice Model to the Aggregate model for Location 8	98
35	Comparing the Choice Model to the Aggregate model for Location 5	101
36	Site characteristics of the compressed cloverleaf along Interstate 5 in Burbank, CA ..	106
37	Improvement in ramp illumination following rehabilitation at Location 6	108

LIST OF TABLES

Table	Description	Page
1	Optimal parameters and associated Log-Likelihoods for Location 8	47
2	Optimal parameters and associated Log-Likelihoods for Location 5 (Yield Line)	64
3	Optimal parameters and Log-Likelihoods for Yield Line with adjusted error terms. . .	69
4	Combined Log-Likelihoods for the Larger Sign and Yield Line data	75
5	Optimal parameters and Log-Likelihoods for Yield Line based on Unified solution . .	80
6	Second Iteration Combined Log-Likelihoods for Larger Sign and Yield Line	83
7	MOE Likelihood Ratio Test p-values for Larger Sign, using Unified solution	84
8	MOE Likelihood Ratio Test p-values for Yield Line, using Unified solution	85
9	MOE Likelihood Ratio Test results for Larger Sign, using alternate Unified solution .	88
10	MOE Likelihood Ratio Test results for Yield Line, using alternate Unified solution . .	89
11	Compliance rates broken down by each driver's arrival situation	94

ACKNOWLEDGMENTS

I would like to acknowledge the FHWA Universities and Grants Programs office for supporting this research through the Eisenhower Fellowship Program, without whose funding this project would not have been possible. Furthermore, the Eisenhower Fellowship Program provided the opportunity to attend and present at multiple TRB annual meetings, which played an invaluable role in expanding the depth and breadth of my knowledge as a transportation researcher and aspiring practitioner.

Also playing a crucial role in this research effort was the Caltrans District 7 (Los Angeles) Office of Traffic Investigation, and Mr. Sheik Moinuddin, Mr. Joseph Kibe, and Mr. Norman Hui in particular, for their help with getting compliance strategies implemented in the field. It was by their efforts that this thesis went from a concept on paper to a reality in the field, and for this I am deeply indebted to them.

Furthermore, I would like to thank my thesis committee chair, research advisor, and friend Dr. Alex Skabardonis, for always being there to guide me and my efforts, for always encouraging me to pursue my research and professional goals whatever they were, and for helping me find solutions whenever I hit problems that I could not solve on my own. I also want to thank the other members of my thesis committee—Professors Robert Leachman and Samer Madanat—for always making time to provide feedback and guidance, for granting me the freedom and flexibility to carry out this research and search for solutions independently, and for being supportive and respectful through everything.

When I was still an undergraduate at UCLA, Mr. Eric Shen—a friend and mentor from the Port of Long Beach—saw great potential in me and encouraged me to pursue a Ph.D. and to apply for an Eisenhower Fellowship. Although I have already acknowledged that this dissertation would not have been possible without the support of the Eisenhower Fellowship Program and the guidance of my committee, I never would have had either of those things to begin with if not for Mr. Shen's encouragement, mentorship, and support.

I am also very grateful to the following families for their years of support, hospitality, and encouragement: Mr. Miguel Montaña & family, Ms. Anne Le & family, and Mr. Art Caire & family. These three families showed concern for my equipment, my research, and for me as a friend, and I was deeply touched by how willingly they welcomed me into their lives. Without their help for the three years that I conducted my research, I would not have been able to collect the on-ramp traffic data necessary for this dissertation.

The volume of video data associated with this dissertation was daunting, but with the help of several research assistants, the data processing took far less time than I could ever have done on my own. Thus, I also want to thank those who made it possible to complete this dissertation in a timely manner through their data processing contributions: Apaar Bansal, Kevin Campbell, Nathan Chan, Emily Chu, Jonathan Lee, Sandra Lee, Katie Leung, Jordon Toy, Alan Joseph Uy, Marcella Wong, and Marlissa Wong. Marcella also gave me continual support and encouragement throughout the analysis stage of my research, and for that I am also very grateful.

Finally, and most significantly, none of this would have been possible without the lifelong support of my loving parents, Susan and Scott, and all the energy, time, and resources they invested in me to make all of this possible. They have given me so much and never asked for anything in return, and whenever I needed anything, they were ready, willing, and eager to do anything and everything they could. Although there are many people that deserve

acknowledgment and thanks in connection with this dissertation, I am by far the most deeply indebted to my mom and dad for everything they have done for me.

1. INTRODUCTION

Noncompliance with traffic controls is a common problem, and agencies are frequently developing new strategies to combat it, including driver speed feedback signs (1), LED highlighting on traffic signs and pavement markings (2), and other forms of innovative signage and pavement markings (3, 4). However, when evaluating the effectiveness of these strategies, researchers generally use conventional before-and-after studies, which are oversimplified and can misrepresent the true effectiveness of the strategies tested. In the “Literature Review” section, we examine the causes and consequences of these shortcomings in greater detail.

To overcome these problems and provide more accurate measures of effectiveness based on the long-term degree of improvement achieved by various compliance strategies, we propose a new framework for analyzing and comparing such compliance measures in the “Methodology” section. As part of the framework, we develop a model to describe drivers’ behavioral response (in terms of compliance) following the implementation of a compliance strategy. This driver response model takes the well documented Novelty effect and combines it with a delayed driver awareness effect to account for the fact that it may take drivers some time to become aware of the change or to understand how to respond appropriately to it. The resultant model predicts a period of improving compliance immediately following the implementation of a strategy, which contrasts with the conventional assumption in literature of a driver response that starts off strong and tapers off over time.

To illustrate the application of this framework and to validate the trend of increasing compliance predicted by our model for the period immediately following the implementation of a compliance strategy, we will use data from a set of yield-controlled freeway entrance ramps along Interstate 10 in eastern Los Angeles. The “Data Collection and Processing” section describes the driver compliance problem in detail, summarizes the data collected as part of the field study, and explains the methods used to quantify driver compliance from week to week during the study period.

In the subsequent section, “Applying the Analysis Framework,” we will analyze the effectiveness of two compliance strategies that were implemented at different entrance ramp locations on the study corridor, using our framework to evaluate the short-term and long-term improvements afforded by each one:

1. Increasing the size of the Yield sign from 36 inches across to 48 inches across, resulting in an increase of the viewable area by 78%.
2. Adding symbolic pavement markings (i.e., a Yield Line) to supplement the existing yield sign.

For each strategy, we need to be able to predict what the driver compliance rates would have been from week to week if nothing had changed at the experimental location, so that we can reliably isolate the improvement in the post-change weeks due to the compliance strategy (and not due to other external factors). To accomplish this, we will identify the model that best predicts driver behavior at each of the experimental locations for the pre-change weeks based on a set of reasonable covariates—for example, the observed compliance rates at other nearby locations. As part of our analyses, we will also be evaluating the durations of the Novelty effect and the delayed driver awareness effect for each compliance strategy, which will allow us to distinguish between the short-term effects and the long-term improvements.

Following the separate analyses of the Yield Line and Larger Sign, we will then merge the findings into a single, consistent set of unified results in the section titled, “Unified Results, Sensitivity Analysis, and Alternate Approaches.” Because this unification was found to support

two related, but distinct, conclusions regarding two of the model parameters, we explore both outcomes separately as part of a sensitivity analysis and find that the overall measure of effectiveness for each compliance strategy changes by only a small margin when switching between the two outcomes. We also establish 95% confidence intervals for our measures of effectiveness, and again find that our choice of unified results has only a minor effect on these intervals.

Later in the same section, we also examine three other forms of our prediction models for the experimental locations (i.e., the models that predict what the driver compliance rates would have been like if nothing had changed) and confirm that the ones we have used in our analysis are the most accurate ones available. We finish by comparing our measures of effectiveness for the larger Yield sign and the Yield Line to those that would have been obtained from a more conventional before-and-after approach, to illustrate the consequences of failing to properly account for delayed driver awareness and learning effects, the Novelty effect, and other external factors that may affect driver behavior during the post-change evaluation period.

Over the course of this three-year study, two additional experimental changes also occurred that may be expected to have an effect on driver yielding behavior: the installation of supplemental pavement lettering (i.e., the word “Yield” on the pavement) at one location, and the improvement of lighting to enhance visibility at night at another location. These two changes represent immediate future research opportunities for applying our framework. In the “Discussion” section at the end of this dissertation, we explore other potential applications as well, and summarize the findings and contributions of our work.

Outline

The following summarizes the overall organization of this dissertation:

- Examine current practices (and their limitations) for evaluating compliance strategies.
- Develop new framework that overcomes these issues.
- Show how framework may be applied to evaluate larger signage.
- Apply the framework to additional cases, and combine these findings with those for the larger signage to arrive at a set of consistent, unified results.
- Demonstrate benefits of this framework compared to traditional before-and-after studies.
- Discuss applications and future work.

2. LITERATURE REVIEW

When using before-and-after data, the typical approach is to simply use the “before” period to establish a baseline for comparing the “after” data to (3, 4, 5). However, this approach assumes that conditions remain unchanged between the “before” and “after” period, which is often not the case. Even when a control group is included to address such changes that may affect driver behavior, we must make the assumption that conditions are identical between the control site(s) and the experimental site(s) at all times—an assumption that can be untenable if, for example, congested conditions occur more frequently or are more severe at the experimental site(s) than at the control site(s). A better approach would minimize our reliance on this assumption by using the “before” data to develop a predictive model for driver behavior at the experimental site, based on statistically significant covariates (e.g., prevailing weather conditions, traffic volumes).

Control groups are also necessary for a sound analysis of any compliance strategy, to account for other factors not explicitly included in the prediction model that may affect driver behavior apart from the experimental change, such as seasonal traffic pattern variations, amount of daylight present (which will vary throughout the year), and even the possible psychological effects of holidays or recent Daylight Saving Time adjustments. The behavior of drivers at relevant control sites would become additional candidate covariates for our prediction model, and would be included if they proved to be statistically significant to the model.

For a control group (or control site) to be “relevant,” its design and environmental conditions would need to be similar to that of the experimental site (or subject site), but it would need to draw upon a different pool of drivers so that the effects of the experimental change could be isolated to the experimental site. Furthermore, the availability of several sites of similar configuration, design, and conditions enables us to test multiple compliance strategies, so that we may subsequently make comparisons of effectiveness among them. Trying to test multiple strategies using only a single site would not be appropriate, as the behavior of the site’s regular users would be affected in an irreversible way following the implementation of the first change. That is, it would not be possible to simply “reset” the drivers to their original state by reverting the site to its original condition due to drivers’ memories—this would consequently have an impact on the observed measures of effectiveness for all subsequently implemented strategies.

The difficulty associated with finding such a cluster of similarly designed sites is likely a leading cause for the frequent absence of control groups in compliance strategy analyses (1, 3). In some cases (e.g., evaluation of strategies dealing with speed limit compliance), upstream or downstream portions of the same roadway are used as controls for the experiment (4, 5). However, differences in geometry (i.e., roadway grade, width, curvature, etc.) between the “control” area of the roadway and the “experimental” area limit the appropriateness of this kind of approach.

Because many real-world studies focus on just one site, they typically involve only one tested strategy and do not include any inherent comparisons of effectiveness to others (3, 5, 6, 7). Furthermore, it is difficult to make such comparisons using the results from other before/after studies due to the unique details of each study site chosen. One existing study attempts to compare the effectiveness of multiple strategies by implementing different ones at different sites in the same region, but this paper suffers from the problem of significant geometric differences between study locations (4).

To overcome these hurdles associated with proper control groups, some researchers rely on driving simulators in place of real-world data (6, 7). However, these studies involve subjects who do not have the same mental state as drivers on the road, as the subjects know that their behavior is being monitored and are fully focused on the task at hand (i.e., their thoughts are not likely to be on other things during the driving simulation). Some simulators cannot provide realistic feedback (such as vehicle noise or vibrations), nor can they accurately reproduce the mentally demanding complex visual environment that drivers often face.

Another approach researchers have come up with is to place test subjects behind the wheel of a real vehicle that has been instrumented to record information (such as speed and acceleration) (5). However, as these participants are presented with a vehicle they are not accustomed to driving, and are aware that their actions are being constantly monitored, they are likely in a state of heightened awareness during the experiment which is not reflective of real-world driving. Furthermore, although subjects are told to “drive the way they normally do,” there is the possibility that they will alter their behavior out of a desire to be good drivers for the

experimenters, or a subconscious psychological desire to try to give the researchers the results they want.

Another frequent limitation of existing studies based on real-world data is that the data are aggregated into three large periods: a “before” period, an “acclimation” or novelty period, and an “after” or steady state period (3, 4). This is an unavoidable outcome in some cases, as some studies rely on collision data as a measure of effectiveness, and such collisions are so infrequent that aggregation is necessary to achieve reasonable sample sizes. As a consequence of aggregation, the details of what happens in each of these periods—in particular, the acclimation period—are lost. Researchers understand that the effect is unstable in the “acclimation” period and stable in the “after” period, but the lack of investigation into driver behavior during the acclimation period means that there is also uncertainty regarding how it changes over time within this acclimation period (though conventional explanations are that it starts strong and tapers off) (3, 4) or how long the period lasts, with some studies suggesting a period of 2-3 weeks for driver acclimation (4) and other studies using a period of as many as 7 weeks (time period chosen in advance, rather than in response to data that was gathered) (3). Studies based on simulated environments tend to disregard the effect of driver acclimation/novelty altogether, since they typically rely only on single trial runs (done in one sitting) for each participant (5, 6, 7).

Finally, unlike the strategies we will be testing in our study, many of the compliance strategies examined in the current literature are unique to the particular problem they are trying to address, and are not generalizable to other situations where noncompliance is a problem (5, 6).

3. METHODOLOGY

To address the limitations of the evaluation methods currently used in the literature, we propose and test a new framework for more accurately measuring the effectiveness of driver compliance strategies.

Site selection

First, an experimental site must be chosen that has other nearby sites of similar configuration that may be used as control sites. Examples include a set of several adjacent parallel streets, intersections along an arterial, or exit ramps from a freeway. Care must be taken to ensure that the different sites are comparable with respect to land uses, sight distance, alignment with the sun, and any other factors that one would expect to affect driver behavior. Furthermore, the drivers of interest should not be captured in the data for more than one site, since the effects of the experimental change could then be observed in their subsequent behavior when those drivers traverse the control sites as well.

For example, if we considered a strategy dealing with driver yielding at pedestrian crosswalks, we would not want the crosswalks to all be along the same arterial, because any strategy implemented at one “experimental” crossing may result in increased driver awareness of other “control” crosswalks along the arterial as well, resulting in observed improved behavior at these control sites that would make the strategy seem less effective than it really is. Even if the experimental crosswalk were downstream of the control crosswalks such that it were the last thing users were exposed to, the analysis would still be skewed by any repeat users of the site who, upon traversing the corridor a second time, might have a newfound awareness of all the pedestrian crossings along the corridor including the control ones. A proper experimental design would look at a set of crosswalks on several parallel streets, since we would expect the drivers on

each street to be coming from a different user pool (i.e., someone traveling north on 1st Street would not also be expected to travel north on 2nd Street in general).

Creating a Driver Behavior Prediction Model

Having selected appropriate experimental and control sites, we then collect a set of “before” data to establish a predictive model for the usual behavior of drivers at the experimental site(s). We split the “before” period into several bins representing equivalent spans of time from one day or week to another, based on the temporal resolution we want for our “after” analysis. For example, we might select a weekly resolution for a compliance strategy whose effect we expect will take months to stabilize, while we might select a daily resolution for a compliance strategy whose effect we expect will stabilize in just a couple weeks. Selecting a resolution that is too high may make the analysis excessively time-consuming to carry out, whereas a resolution that is too low will mask the subtleties of the short-term driver novelty response (3, 4). The bins should also be consistent with respect to the time(s) of day and day(s) of the week captured. For example, if using weekly bins, we might consider the mid-day period from 10 AM to 2 PM on Monday through Friday.

For each bin, we then evaluate the driver compliance rate at each experimental and control site, as well as other environmental variables that we would reasonably expect to have an effect on driver compliance such as average traffic density, predominant Level of Service conditions, prevalence of sunlight or ambient light, and average visibility. Many derivatives of the above factors are also possible, such as the percent of time the speed falls below a certain threshold for the bin, or the natural logarithm of the average speed, and it is up to the researchers’ judgment to decide which variables are most sensible candidate covariates for the particular compliance problem being examined.

Once the dependent variable (i.e., the compliance rate at the experimental site) and a set of potential covariates are all evaluated for each bin in the “before” period, we test different combinations of these variables in a generalized linear model where we apply a Logit-Logistic transformation to the dependent variable to ensure that the model always predicts a compliance rate between 0 and 1. Depending on the nature of the compliance strategy being evaluated, it may be reasonable to restrict the model to variable combinations that do not simultaneously include two or more inherently correlated variables (e.g., “average freeway speed” and “percent of time the speed is below 35 mph”), though this is left to the judgment of the researchers.

In developing our model to predict behavior at the experimental site using a subset of the candidate covariates described earlier, our goal should be to minimize the prediction risk, not to maximize the goodness of fit to the training data set (i.e., to the “before” data). Although the model with the best predictive power will typically have a high R-squared value as well (relative to other model formulations), it is not sufficient to maximize R-squared, since this model will not necessarily have the highest predictive power. For example, R-squared can easily be inflated by adding new variables that do not contribute significantly to the model.

Using a cross validation technique such as 10-Fold Cross Validation (8) with a squared error loss function, we can identify the best model as the one that minimizes prediction risk. In doing so, we protect against selecting a model that has been over-fitted to the “before” data but has sacrificed prediction accuracy over a new dataset in the process. This technique reserves 10% of the data for validation, and uses the remaining 90% to solve for the model coefficients. The model is then applied to predict the results of the validation dataset, and the deviation used

to estimate the prediction risk associated with the model. The procedure is then repeated 10 times, with the data rotated so that all of it is included in the training set at some point.

A paired one-tailed t-test can be used to evaluate whether one model is significantly better at predictions than another based on the risk associated with each of the 10 “folds” for each of the two models being compared. For example, such a test can provide a p-value on the hypothesis that a particular model formulation is no better than a constant-only model. The Durbin-Watson statistic can be used to check for serial correlation, which can then be addressed if necessary using one of several common techniques (9).

Once we identify the model formulation that provides the most accurate predictions, we use this model along with its covariates to predict the behavior of the experimental location in the period after the compliance strategy is implemented. We will refer to this *baseline compliance prediction model* as $P(t)$, which predicts compliance at the experimental site at time t if nothing had changed at $t = 0$. From this, we can define the baseline noncompliance prediction model as $N(t) = 1 - P(t)$. Confidence intervals for the model predictions can be estimated using the mean and standard deviation (σ_{cv}) of the errors from the cross-validation.

Comparing Predicted and Observed Driver Response

The objective of a compliance strategy is to reduce the frequency of non-compliant drivers, suggesting that our measure of improvement for a particular strategy should be the percent reduction in violation rate. A strategy that makes no difference in driver behavior results in a 0% improvement, and a strategy that results in perfect compliance has a 100% improvement. For a given time t after the compliance strategy is implemented, improvement is defined as:

$$Improvement(t) [\%] = \frac{N(t) - O(t)}{N(t)} \quad (1)$$

where $O(t)$ is the observed noncompliance rate at the experimental site at time (t). The above formula can be used to evaluate the degree of improvement realized in each bin of the “after” period. This binned data on degree of improvement (by days or weeks since the strategy was implemented) can then be modeled by a fitted *Driver Response Curve*, whose asymptotic behavior represents the long-term degree of improvement resulting from the experimental change—and our overall measure of effectiveness for the compliance strategy tested.

Modeling the Driver Response over Time

The driver response curve is the combination of two effects, which we will call the *Awareness Effect* and the *Novelty Persistence Effect* (Figure 1). Depending on the nature of the compliance strategy, it may take drivers some time to notice and properly interpret it. For example, installing an oversized sign may be subtle enough to go unnoticed for a while, whereas adding LEDs to the original sign may be recognized immediately after being installed. Alternately, writing the word “Yield” on the pavement may be immediately understood, whereas drivers may need to be exposed to a Yield Line several times before they ascertain what it means.

The Awareness Effect captures the delay between (1) installation and (2) driver awareness and understanding of the strategy’s presence (Figure 1a). We might intuitively expect this Awareness Curve to follow a Normal distribution, a Geometric distribution (i.e., each time drivers traverse the site, there is a set probability that they will notice the change if they have not already), or possibly a simpler Uniform distribution with a linear CDF (i.e., roughly the same

(a) Drivers' awareness of the change

The above case assumes a Normal Distribution for how long it takes drivers to notice a change in their driving environment.

(b) Persistence of Novelty Effect

The above case assumes a Uniform Distribution for the persistence of novelty among different drivers. The Novelty Effect does not affect all drivers, and only occurs after drivers have noticed the change according to (a).

(c) Observed short-term Novelty Response

The shape of this curve is determined by the shapes of the Driver Awareness curve (a) and the Persistence of Novelty Effect curve (b).

(d) Observed long-term response

This represents the Driver Awareness effect (a) when the Novelty Effect drivers are excluded.

FIGURE 1: Examples of distributions for (a) the Driver Awareness effect and (b) the Persistence of Novelty Effect, and the resultant driver response in terms of (c) short term Novelty Effect compliance and (d) long-term compliance.

number of unaware drivers notice the change for the first time in each time interval). The Uniform distribution has an undesirable kink point in the CDF when awareness hits 100%, however, and it is because of this characteristic that the Uniform model seems relatively unrealistic compared to the alternatives. Literature suggests that the time it takes drivers to become aware of and respond to a change in their driving environment follows a Log-Normal distribution, with a standard deviation, σ_a , equal to approximately half of the mean, μ_a (10, 11, 12). However, it may not be appropriate to attempt to generalize these results to our current situation involving repeat exposure to environmental elements that may go unnoticed for several visits, so researchers may want to test other statistical distributions as well (e.g., Geometric, Uniform, Normal) and select the one that yields a driver response curve that best fits their improvement data for the “after” period.

The improvement (i.e., reduction in violation rate) observed at any given time after the strategy has been implemented is made up of those formerly noncompliant drivers who noticed the change and are compliant (1) from that moment onward, or (2) as long as the novelty effect of the strategy persists. **Group 1** represents a long-term improvement, and captures those drivers who were formerly noncompliant simply because they were unaware of what they were supposed to do (e.g., had never noticed an unmarked crosswalk until now). **Group 2** represents a short-term improvement that will wear off once the novelty of the compliance strategy (e.g., novel striping for speed humps) fades and those drivers become acclimated to the change. We would expect Group 2 to generally be smaller for a compliance strategy that involves enforcement, since these drivers are essentially aware of the proper behavior and are simply choosing not to obey after a while. We will use β_a to represent the total proportion of originally non-compliant drivers captured by Groups 1 and 2. Of this proportion of drivers who exhibit newly compliant behavior, either permanent (i.e., Group 1) or temporary (i.e., Group 2), we will use β_n to represent the fraction that are in Group 2. Note that we make the assumption that the factors that cause $N(t)$ to vary over time affect the all categories of non-compliant drivers comparably, such that β_a and β_n are not functions of t .

We will refer to the drivers who are noncompliant and who never notice or respond to the change (i.e., noncompliant drivers that are not in Group 1 or Group 2) as members of **Group 3**. The fraction of drivers in this group is given by:

$$\alpha_3 = N(t) \cdot (1 - \beta_a)$$

The fraction of drivers who are noncompliant but eventually notice the change and become compliant from that moment onward (i.e., Group 1) is given by:

$$\alpha_1 = N(t) \cdot \beta_a(1 - \beta_n)$$

The fraction of drivers who are noncompliant, eventually notice the change, and respond to the novelty of it by complying until its novelty fades away (i.e., Group 2) is given by:

$$\alpha_2 = N(t) \cdot \beta_a(\beta_n)$$

Summing α_1 , α_2 , and α_3 gives us $N(t)$ as expected, representing all noncompliant drivers combined. The overall compliance rate $C(t)$ at a given time (t) after the strategy has been implemented is modeled by:

$$C(t) = P(t) + R(t) + S(t) \quad (2)$$

where:

$P(t)$ is the predicted compliance at the experimental site at time (t) if nothing had changed.

$R(t)$ is the additional improvement in compliance rate at time (t) due to Group 1 drivers

$S(t)$ is the additional improvement in compliance rate at time (t) due to Group 2 drivers.

The $P(t)$ term was described earlier, in the section titled “Creating a Driver Behavior Prediction Model.” The $R(t)$ effect is rather straightforward: they become compliant according to the shape of the Cumulative Density Function $F_a(t)$ for the Awareness Curve, since the only factor affecting their response is how long it takes for these drivers to notice the change (Figure 1d).

$$R(t) = \alpha_1 \int_{t_a=0}^{t_a=t} f_a(t_a) dt_a = \alpha_1 F_a(t) \quad (3)$$

Where $f_a(t_a)$ is the probability density function of the Awareness Curve. Recall that literature suggests this curve follows a Log-Normal Distribution with $\sigma_a \approx 0.5\mu_a$. $F_a(t)$ is a cumulative density function representing the fraction of Group 1 drivers who are expected to have noticed (and reacted to) the change by time t . Note that we make the simplifying assumption that the Awareness Function $f_a(t_a)$ does not differ between Groups 1 and 2.

The $S(t)$ effect is complicated by the fact that it depends on both how long it took for a given driver to notice the change, and how long the novelty of the change persists once observed, since the novelty effect cannot take effect until after a given driver has noticed the change.

$$S(t) = \alpha_2 \int_{t_a=0}^{t_a=t} [f_a(t_a) \cdot Q(t_a)] dt_a \quad (4)$$

The integral in Equation 4 represents the fraction of Group 2 drivers whose novelty effect is still active at time t , where $Q(t_a)$ is defined as:

$$Q(t_a) = 1 - \int_{t_n=0}^{t_n=t-t_a} f_n(t_n) dt_n \quad (5)$$

Here, $f_n(t_n)$ is the *Novelty Persistence* probability density function, which describes how long it takes for Group 2 drivers to revert to their former noncompliant behavior once they have noticed the compliance strategy (Figure 1b). Thus, the integral in Equation 5 measures what fraction of those Group 2 drivers that became aware of the change at time t_a have lost their novelty response by time t because the persistence of the effect for these drivers was shorter than $t - t_a$. We then integrate across all values of t_a from 0 to t in Equation 4 to get the overall improvement in compliance at time t due to the contribution of the Novelty Effect (Figure 1c).

Note that for $f_n(t_n)$ and $f_a(t_a)$, any nonzero parts of the probability distribution function that occur over the negative portion of the time domain are instead ascribed to the time $t = 0$, which ensures that we do not improperly clip or neglect any parts of either distribution. To ensure reasonable results, we will restrict ourselves to forms of $f_n(t_n)$ and $f_a(t_a)$ that have only a minimal probability over the negative portion of the time domain in our analyses.

Novelty Persistence Curve

There is scarce literature dealing with the distribution of Novelty Persistence $f_n(t_n)$, despite how well documented the Novelty Effect is (1, 3, 4, 13). We might intuitively expect the distribution to follow a Normal curve profile (indicating that the amount of time something remains new and exciting to drivers is a Normally distributed parameter), or possibly a Geometric distribution (suggesting that with each exposure to the new item, there is a certain probability that a driver will no longer find it interesting). However, more research is needed in this area, and in the meantime researchers may want to test other reasonable distributions as well (e.g., Uniform) and select the one that leads to a driver response curve that best fits their improvement data for the “after” period. Our results presented later suggest that the persistence of the Novelty Effect follows a Uniform or a Normal Distribution.

Although the Novelty Persistence curve can be expected to have the same distribution type for different situations, we would not expect the shape parameters themselves to always be the same. In other words, if the Novelty Persistence curve for *Strategy A* follows the shape of a Normal distribution, we expect the Novelty Persistence curve for *Strategy B* to also follow a Normal distribution, but would not expect it to have the same mean or standard deviation. We can estimate the standard deviation and mean for a strategy’s novelty persistence through curve fitting to see which parameters allow for the best fit between the predicted and observed behavior at the experimental location.

However, we can reduce the number of degrees of freedom (i.e., curve-fitting parameters in our model) by one if we assume that the standard deviation for the Novelty Persistence curve scales proportionally with the mean, just as we assumed for the Awareness Curve. For example, if the mean of *Strategy A*’s Novelty Persistence curve is twice that of the mean for *Strategy B*, we assume the standard deviation is twice as large as well. This can be justified by realizing that if (for example) the mean for *Strategy A* is four weeks and the mean for *Strategy B* is two weeks, it follows that *Strategy A* remains novel to drivers for about twice as long on average. Those drivers who lost interest twice as fast as the mean for *Strategy A* could be expected to lose interest twice as fast for *Strategy B* as well. Applying this logic to all drivers in this example (i.e., drivers at each percentile, and not just at the mean) justifies our assumption of a standard deviation that scales proportionally with the mean, since a mean that is twice as large implies an effect that lasts twice as long on average. This has the consequence of scaling the Novelty Curve horizontally by a factor of two, which increases the standard deviation by a factor of two as well.

Measure of Effectiveness (MOE)

Because our objective is to minimize the size of the group of noncompliant drivers, our measure of effectiveness should capture how well our strategy accomplishes this in the long run, when the driver response has stabilized. The two effects that reduce the degree of noncompliance $N(t)$ in a given week are the long-term *Awareness Effect*, represented by $R(t)$, and the short-term *Novelty*

Effect, represented by $S(t)$. As $t \rightarrow \infty$, $R(t) \rightarrow \alpha_1$ and $S(t) \rightarrow 0$, so our measure of effectiveness (MOE) is:

$$\text{Effectiveness} = \frac{\alpha_1}{N(t)}$$

In this way, a perfect strategy will reduce the number of violators by 100% (making it 100-percent effective), and a completely ineffective long-term strategy will reduce it by 0%.

Curve-Fitting Parameters

The overall Driver Response Curve (Equation 2) returns the expected compliance rate by week, when adding in the effects of both Novelty and Driver Awareness. The shape of this curve is fully defined by $N(t)$, β_a , β_n , $f_a(t_a)$, and $f_n(t_n)$. $N(t)$ is evaluated from the “before” data, and the remaining terms are adjusted to find the values that result in the best fit between the predictions from Equation 2 and the observed data for those weeks, based on likelihood maximization. As explained earlier, we may make assumptions about the distributions of $f_a(t_a)$ and $f_n(t_n)$, or we may test different types of standard distributions and select the one (along with its shape parameters) that provides the best fit.

Once we have solved for all the shape parameters that define Equation 1, we can establish confidence intervals for our MOE by constraining α_1 to different values and applying a likelihood ratio test. We may also test the significance of the Novelty Effect and Driver Awareness analogously, by constraining either α_1 or α_2 to zero.

Before performing any such significance tests or likelihood maximizations, however, we must convert all deviations between the observed and predicted values of each week (assuming the errors are normally distributed) into Z-scores for a standard Normal distribution using the Error Term equations in the next section to estimate the Standard Deviation for each point. This is due to the fact that the continuous version of the Likelihood Function requires all observed points to come from the same probability density function.

Error Term

The error term associated with the model predictions of Equation 2 are based on the errors associated with the uncertainty σ_{cv} in the baseline compliance prediction model, $P(t)$. All other terms are shape parameters for the Driver Response Curve and are determined by fitting the model to the observed data: β_a , β_n , $f_a(t_a)$, and $f_n(t_n)$. When we expand Equation 2 and take the partial derivative with respect to $P(t)$, we obtain:

$$dC(t) = \left[1 - \beta_a F_a(t) + \beta_a \beta_n F_a(t) - \beta_a \beta_n \int_{t_a=0}^{t_a=t} [f_a(t_a) \cdot Q(t_a)] dt_a \right] dP(t)$$

Thus, the approximate error in our model predictions of Equation 2 is:

$$\sigma_c = \left[1 - \beta_a (1 - \beta_n) F_a(t) - \beta_a \beta_n \int_{t_a=0}^{t_a=t} [f_a(t_a) \cdot Q(t_a)] dt_a \right] \sigma_{cv}$$

Adjustments for Compliance Strategies that do not address enforcement issues

When dealing with a compliance strategy that does not relate to enforcement, it may be more realistic to assume that some fraction of drivers is being deliberately noncompliant. Furthermore, we assume that this fraction, which includes Group 2 drivers, will not change during the study period apart from any temporary novelty effects, given there are no changes in levels of enforcement. Whether this assumption is appropriate or not will depend on the nature of the study and is left to the judgment of the researchers. In this section, we develop a slightly modified version of our original framework that more properly characterizes the effectiveness of a strategy that is not intended to deal with any underlying enforcement problems.

We begin by specifying in advance the size of this noncompliant group as a constant fraction β'_n of all drivers. For this framework, we must have prior knowledge about β'_n . For instance, we might estimate it from other data (e.g., the observed decrease in violations when visible enforcement is present). We then define β'_2 as the fraction of this noncompliant group that exhibit a temporary novelty response (i.e., the fraction of β'_n that belongs to Group 2), and α_2 becomes:

$$\alpha_2 = \beta'_n \beta'_2$$

Therefore, $N(t) - \beta'_n$ represents the members of Groups 1 and 3. We define β'_a as the fraction of these remaining noncompliant drivers that belong to Group 1, such that:

$$\alpha_3 = (1 - \beta'_a)(N(t) - \beta'_n)$$

$$\alpha_1 = \beta'_a(N(t) - \beta'_n)$$

Because we are specifically considering compliance strategies here that do not address enforcement problems, they cannot be expected to capture any of the deliberately noncompliant drivers in β'_n in the long term. Instead, our MOE should measure how the size of Group 1 compares to the total number of noncompliant drivers excluding β'_n :

$$\text{Effectiveness} = \frac{\alpha_1}{N(t) - \beta'_n} = \beta'_a \quad (6)$$

When maximizing the goodness of fit of our Driver Response Curve to the observed data, we have the same set of shape parameters except that β_a and β_n are replaced by β'_a and β'_2 . We also have the added constraint that $\beta'_n < N(t) \forall t$, since the number of deliberately noncompliant drivers cannot exceed the total number of noncompliant drivers at any time. If $N(t)$ is lower than β'_n , we must lower β'_n to this new value so that the inequality is not violated, and then repeat our analysis. However, this constraint would normally be expected to remain inactive, as setting $\beta'_n = N(t)$ results in the unusual situation of a Group 2 size equal to zero and leaves no room for long-term improvement in the absence of enforcement.

Once we have solved for all the shape parameters that define Equation 2, we can establish confidence intervals for our MOE by constraining α_1 to different values and applying the likelihood ratio test, which is equivalent to constraining β'_a . We may also test the significance of the Novelty Effect and Driver Awareness analogously, by constraining either β'_a or β'_2 to zero.

Our new formula for the approximate error associated with the model predictions of Equation 2 when using the definitions of α_1 , α_2 , and α_3 from this subsection is:

$$\begin{aligned}\sigma_c &= \sqrt{(\sigma_{c,cv})^2 + (\sigma_{c,\beta'_n})^2} \\ \sigma_{c,cv} &= [1 - \beta'_a F_a(t)] \sigma_{cv} \\ \sigma_{c,\beta'_n} &= \left[\beta'_2 \int_{t_a=0}^{t_a=t} [f_a(t_a) \cdot Q(t_a)] dt_a - \beta'_a F_a(t) \right] \sigma_{\beta'_n}\end{aligned}\tag{7}$$

Where $\sigma_{c,cv}$ is the error in the predictions of Equation 2 due to the uncertainty σ_{cv} in the baseline compliance prediction model, $P(t)$, and σ_{c,β'_n} is the error in the predictions of Equation 2 due to the uncertainty $\sigma_{\beta'_n}$ in the estimate for β'_n .

4. DATA COLLECTION AND PROCESSING

To illustrate the application of this framework, we examine a compliance problem associated with yield-controlled merges on freeway entrance ramps to Interstate 10 in Los Angeles. For this analysis, we rely on video data to monitor driver yielding behavior, using a set of carefully defined criteria to evaluate whether a given driver yields properly or fails to yield when conditions warrant it. After the video data is used to characterize the extent of the compliance problem, different compliance strategies will be implemented at a subset of the locations. Using our framework, we will evaluate the effectiveness of each one, and will draw comparisons about the relative benefits of each strategy.

Background

When the San Bernardino Freeway (Interstate 10 in Southern California) was being constructed, a large portion of its routing in the San Gabriel Valley (east of downtown LA) was adopted from the former Ramona Parkway, an expressway with lower design speeds. The mainline did not suffer much from this repurposing of old right-of-way, since the layout is almost entirely straight and flat as it passes through the San Gabriel Valley. However, the street interchanges—built as compressed cloverleaf interchanges (see Figure 2)—were not as easily adapted to the faster-moving traffic or the higher vehicle volumes, as they require very sudden deceleration and acceleration of entering and exiting traffic, and have such tight ramp merges that Yield signs are needed to stop vehicles on one approach when another approach has oncoming cross-traffic. Without this yielding requirement, entering traffic would have so little time to negotiate for space at the merge point that the situation could result in frequent collisions.

Figure 2 shows a typical interchange for the freeway entrance and exit ramps on this corridor, with entering traffic following one of two paths. The first path (shown in yellow) includes an auxiliary lane stretch that doubles as an acceleration zone, and a short merge onto the freeway farther downstream. The other path (shown in red) includes a yield point and subsequent tight merge where the two streams join. At this yield point on the red path, there is a standard Yield sign 36" wide (measured along the top) with no supplemental pavement markings.

Figure 3 reveals the suitability of this corridor for testing a compliance strategy, as there are several sites all in close proximity (the highlighted corridor is roughly 4 miles), with a common design but different user pools at each location. The freeway has a roughly constant

FIGURE 2: Typical freeway-arterial interchange design.

FIGURE 3: Corridor layout and prevalence of compliance problem.

grade and width throughout, and maintains a straight alignment. The surrounding land uses are predominately low-density residential, with local commercial destinations along some of the arterials but no large regional commercial destinations except at Rosemead Blvd (excluded from this analysis) and Atlantic Blvd (used only as a control site, and only on the side of the freeway that does not capture visitors of the nearby regional commercial destination).

Data Source

Although these ramps have inductive loops embedded in them, we cannot use these as data sources due to the fact that red-path drivers do not always yield in the same location, meaning that only some of the proper yields would occur over a loop detector while others would go undetected. Additionally, some of the ramp detectors were not functioning properly at the time of the study, such that no data could be obtained from them at all. For these reasons, we use video data to observe driver yielding behavior.

All data for the subject ramp was processed manually by watching the raw video data and recording all instances of proper yielding and failures to yield. Compliance rates were obtained from video data at each location on Sundays between 11 AM and 2 PM. Sundays were chosen for data collection due to recurring congestion that occurred all other days, which interfered with our analysis since yielding behavior deteriorates into turn-taking at the merge point when the ramps become congested. Although we are only considering the midday period, we have approximately 24 hours of video data available to use between Saturday evening and Sunday evening. Furthermore, all video data is kept in an archive, and no footage is deleted even after it has been processed.

Data Overview

A summary of all the data we have collected and processed is given in Figures 4-6. Gaps in the data are present (e.g., the weeks shown in gray) due to a combination of equipment problems and a perceived lack of need for weekly-resolution data several weeks or months after each compliance strategy was installed. Furthermore, weeks where rain occurred were generally unusable, as droplets would cover the camera lenses, distorting the images and making the video footage difficult or impossible to interpret. In addition, the fact that we had no rainy weather days in the “before” period (i.e., October 18 through Dec 20, 2009) made it difficult to establish any proper controls for how driver behavior would change at each of our experimental ramps in the event of rain, thereby preventing us from making valid predictions for $P(t)$ on rainy days.

As Figures 4-6 also indicate, each of the entrance ramps along this corridor was closed at some point for rehabilitation. This was unrelated to our research effort, and was simply a consequence of Caltrans’ maintenance schedule for the freeway. When each interchange was reopened, a number of superficial changes had occurred: new curbs were poured, deteriorating guard-railing was replaced, landscaping was carried out, the pavement was resurfaced (with the same type of material), and pavement markings were repainted (but not changed from their original positioning). Despite these changes, no adjustments to the ramp layout or design were made: the ramps still had the same alignment, with the same controls in the same general positions (i.e., nothing was added or removed, nor were old signs replaced with new ones). However, the shoulders were made flush with the roadway, rather than being slightly inclined and separated by a shallow gutter as they had originally been. This was found to affect driver behavior, as the more accessible shoulder gave red-path drivers more room (and more time) to

FIGURE 4: Summary of video data collected between Oct 2009 and Sept 2010.

FIGURE 5: Summary of video data collected between Oct 2010 and July 2011.

FIGURE 6: Summary of video data collected between July 2011 and Nov 2012.

try to forcibly cut into the traffic stream. This will be discussed in greater detail when we analyze the effect of the ramp rehabilitations later.

During the periods of closure, traffic that would have otherwise taken that particular entrance ramp was instead detoured to the next one farther downstream on the freeway, guided by construction signage from Caltrans. For example, when Location 3 was closed for rehabilitation in late 2011, this disrupted conditions at Location 1 by introducing a different batch of users to that ramp for the duration of the closure. Furthermore, it increased ramp volumes, leading to an average of 210 events per hour of footage, as compared to 93 events per hour in the four weeks prior to the closure. When Location 5 had to absorb detoured traffic from Del Mar Ave earlier that year, the traffic volumes for the ramp grew from an average of 359 vph in the four weeks prior to the closure to an average of 546 vph while the detour was active (representing an increase of 52%), as shown in Figure 7.

From Figure 7, we also note that once a rehabilitated ramp was reopened and the detour no longer needed, the traffic volumes quickly returned to normal at the ramp that had formerly been accommodating the detoured vehicles. Thus, although we consider the closure of an adjacent upstream ramp to be disruptive to the conditions at a particular location, we assume that the conditions return to normal at that location immediately after the closure is removed upstream. This suggests that drivers affected by the ramp closure were checking it regularly, such that there was little latency between the reopening of the ramp and its return to full utilization by drivers. For example, in Figure 7, the first week shown (Aug 22, 2010) was the first week that Location 5 was open after being closed for two months for rehabilitation. Volumes for that week were already at 321 vph, which was 89% of the average volume carried by Location 5 for the four weeks that followed. Therefore, we make the simplifying assumption that there is no lag between when a ramp is reopened and when drivers begin using the ramp again, so that any Awareness effect related to the usefulness of the wider shoulder (for circumventing the Yield requirement) immediately starts capturing drivers from the entire user pool for that ramp, and not merely from a subset representing just those drivers who have already discovered that the ramp has reopened.

Data Collection Equipment

As Figure 6 shows, data collection ceased at four of the six locations in June 2012. This was due to the expiration of contracts that had been allowing us to record the freeway from adjacent private property (a sample of one such contract is included in Appendix C). We had established these 12-month contracts in June 2009 (with renewals in 2010 and 2011) with local property owners near the freeway, asking for permission to set up our cameras on their land in exchange for compensation so that we could monitor the behavior of drivers (anonymously) on the nearby entrance ramps. This was necessary to avoid long delays to our research effort, which would have occurred if we had tried seeking the appropriate permits from Caltrans to set up our equipment on its property for three years.

Figure 8 shows one of the camera setups used. Two cameras were placed in the box at the top, with a portable power unit at the base. The primary support was a telescoping pole, making it easy to lower the cameras for maintenance and access. These camera stands allowed us to consistently frame the same angle in our cameras from week to week, and at 16 feet, they were tall enough to enable us to view ramps on both sides of the freeway from just one vantage point. Several tie-down cords were used for added stability, and the base of the stand was also anchored into the ground with six 2-foot steel stakes.

FIGURE 7: Entrance ramp volumes at Location 5 before, during, and after closure of the on-ramp immediately upstream at Del Mar Av in 2010.

Equipment to record video at Locations 8 and 9.

Equipment to record video at Locations 1 and 2.

FIGURE 8: Camera stands used to monitor entrance ramp traffic each week from adjacent private property.

Originally, Caltrans had planned on installing the final compliance strategy (the word “Yield” on the pavement at Location 1) in late 2011, such that the expiration of the contracts in June 2012 would have been more than sufficient to capture the driver response in the weeks and months that followed. However, as the installation date continued to get pushed back, we decided at the end of May 2012 that it was more appropriate to let the contracts expire and simply negotiate new ones in the future if needed when the final compliance strategy was finally implemented. When this occurred in August 2012, we decided that given the time, equipment, and budgetary constraints we faced, the best strategy would simply be to record the video footage from a nearby freeway overpass as needed. Therefore, for the weeks after the installation of the “Yield” pavement markings at Location 1, we only have data for the experimental ramp and one control ramp, and only for the period of interest (i.e., from 11 AM to 2 PM) in key weeks.

Defining a Proper Yield

In our assessment of yielding behavior, we will consider all observed instances where yielding would have been appropriate (i.e., a red-path vehicle approaches the merge point when there is traffic on the yellow-path that is close enough to be a hazard and warrant yielding), and then determine how frequently the red-path traffic yielded in these situations. To evaluate whether yielding would be appropriate, we consider the following statement from the MUTCD:

*“The **Yield** sign assigns right-of-way to traffic on certain approaches to an intersection. Vehicles controlled by a **Yield** sign need to slow down or stop when necessary to avoid interfering with conflicting traffic.” (14)*

Thus, for the purposes of this analysis, we will assume that yielding would be an appropriate course of action whenever failing to yield would disrupt the flow of yellow-path vehicles that have the right-of-way. The failure to yield would not necessarily need to create a dangerous situation where two or more vehicles come close to colliding, though it would certainly include such events. Yellow-path traffic may be considered obstructed by merging ramp vehicles if the former must decelerate or brake to maintain comfortable headways with slow-moving vehicles that have entered immediately downstream. In this way, our adopted criteria for assessing whether merging red-path vehicles should have yielded is no different from that used in situations of right-turning vehicles waiting at a red light in the presence of cross-traffic, since it is improper for the turning vehicles to proceed if doing so would cause traffic on the cross-street to decelerate; a collision or near collision is not a sufficient condition for determining that merging vehicles should have yielded.

In a separate study of left-turning traffic, it was found that the smallest acceptable gap size to drivers was just over three seconds (15). Therefore, to evaluate whether a failure to yield had occurred, we used a minimum acceptable gap size of three seconds for red-path traffic. This was found to be an appropriate minimum threshold for red-path traffic to avoid interfering with the flow of yellow-path traffic based on subsequent video analysis of driver behavior on these ramps, due to the alignment of the two ramps and the tendency of yellow-path traffic to use the straight ramp as an acceleration zone for the freeway. For this gap size, failing to yield would result in uncomfortably small headways for the following yellow-path traffic—the resultant headways would be far below 3 seconds, since the entering red-path traffic would not initially be moving at the same speed as the yellow-path traffic. Given that most drivers leave one- to two-

second headways in front of themselves, and observing from video footage that it typically takes the merging ramp vehicles at least another two seconds to accelerate to the prevailing speed of the through-traffic stream, this minimum gap size of three seconds for merging red-path traffic is reasonable. Furthermore, making an assumption about the minimum comfortable following distance for the yellow-path traffic is not a problem, as the red-path traffic would be making the same assumptions when deciding whether to proceed or yield (i.e., red-path vehicles have no way of knowing this comfort distance for a particular yellow-path vehicle, and must assume its value when evaluating whether a gap is large enough for them to fit into).

Because failures to yield often resulted in deceleration of approaching yellow-path vehicles and an artificial inflating of the measured gap size as recorded at the merge point, distance-based gaps were used to approximate the 3-second criteria. For this, 30-40 solitary yellow-path vehicles were sampled at each ramp in the absence of any red path traffic, and a distance of between 100 and 140 feet (depending on the particular ramp) upstream of the merge point was found to correspond to a travel time of three seconds for unobstructed vehicles. For those camera angles that had no distinguishing features at the upstream distance threshold point, graphical overlays were superimposed on the video to facilitate reliable identification of the critical gap size using stationary objects in the frame (i.e., sign posts and light posts) as registration markers.

The yielding zone along the red-path is important to define, as we run the risk of interpreting deceleration in anticipation of the last 90° ramp curve as a yield if we look too far upstream along the red-path. The portion of the red-path where we monitored for yielding behavior was the last 50 feet leading up to the merge point, as this was found to be an approximate maximum distance upstream that ramp traffic would begin to exhibit yielding behavior (i.e., would begin to decelerate as part of a proper yield). Fifty feet upstream of the merge was still within the curved portion, so traffic decelerating in anticipation of the curve would have done so by then. To ensure consistency in data processing between different video clips, we superimposed guide marks over the video image for the red-path traffic as necessary to identify the 50 feet prior to the Yield point, just as we had done for the yellow-path traffic.

Camera Angles and Details

Figures 9 show the camera angles for the six locations we recorded: 1, 2, 5, 6, 8, and 9. These figures have been annotated to show the Red and Yellow paths, and to indicate the distance-based thresholds used in each camera frame. For Locations 5 and 8, software was used to generate overlays on the footage to indicate where the distance-based thresholds were, since there were no elements in the image itself to facilitate the identification of these points. These overlays were necessary at Location 5 only for the weeks after the on-ramp was rehabilitated, due to the fact that several background elements that were originally used to identify the positions of various thresholds for red-path and yellow-path traffic were moved during the ramp closure (i.e., the light poles and the “Do Not Enter” sign in the foreground).

The camera angle at Location 6 was not sufficient to let us view yellow-path traffic up to 3 seconds prior to the merge point, so a smaller footage preview was superimposed in the upper right corner showing the same footage as the main window, but advanced by 1.6 seconds. Doing so enabled easy identification of the moments that yellow-path vehicles were within 3 seconds of the merge point, despite the limited range of the camera angle. One unavoidable consequence of this is that our yellow-path thresholds now become, for this camera angle only, a combination of distance-based and time-based measurements. However, this approximation was made on only

Location 1

Location 2

FIGURE 9 (Part 1 of 3): Camera angles and thresholds used for yielding criteria.

FIGURE 9 (Part 2 of 3): Camera angles and thresholds used for yielding criteria.

Location 8

Location 9

FIGURE 9 (Part 3 of 3): Camera angles and thresholds used for yielding criteria.

this one control ramp, and was applied consistently throughout the entire study period. The preview window was rendered onto the other camera feed, rather than simply overlaid upon it, such that there would be no possibility of the two ever losing synchronization even during fast forwarding, pausing, etc.

Additionally, the main camera angle at Location 9 enabled us to see vehicles as they approached the merge point, but the landscaping next to the on-ramp obscured our view of the merge area itself. To get around this, we were able to set up a convex mirror next to the merge area itself (still on the homeowner's property, not on Caltrans' right-of-way), and recorded the behavior of red-path vehicles at this yield point by aiming a camera at the mirror. This convex mirror was 24 inches in diameter, rated for traffic use, and raised 10 feet high using a modified camera stand to give us a view of the other side of the sound wall that separated the freeway from the private property. The two camera feeds were synchronized using the audio tracks of each recording (since loud noises such as horns honking or dogs barking would have been recorded simultaneously by each camera).

In some cases, when distinctive sounds were not available to aid in synchronization at Location 9, we instead used the travel time of yellow-path vehicles between the right edge of the main camera and the left edge of the mirror camera angle to synchronize the two videos. Specifically, we used the primary camera angle to estimate the speed of a particular yellow-path vehicle, recorded the time that vehicle passed a specific threshold near the right edge of the frame, recorded the time that the same vehicle appeared at a specific spot in the mirror camera frame, assumed the vehicle maintained a constant speed in the travel time between the two video frames (which was well below one second), and calculated an appropriate offset for the two video files using knowledge of the geometry of the ramps and the two camera angles. This process was then repeated several times to obtain a more reliable estimate for the appropriate offset to use.

Theoretical and Practical Considerations

From a theoretical standpoint, it may seem that this distance-based approach does not account for the fact that red-path traffic may arrive at the merge point, notice that an approaching yellow-path vehicle is traveling at such a slow speed that the projected gap size is over three seconds despite the fact that the conflicting vehicle is closer than 120 feet, and proceed through the merge point accordingly. In that case, the red-path vehicle would have been labeled a violator when in fact it was safe to proceed. However, as yellow-path traffic uses this stretch of road to begin accelerating to freeway speeds, we cannot use the upstream speed of yellow-path vehicles to predict how long it will take them to arrive at the merge point—in many cases, vehicles with slower upstream speeds have just begun accelerating, whereas vehicles with higher upstream speeds have already been accelerating for a time. Consequently, their arrival time at the merge point cannot be predicted by their upstream speed alone, or from a combination of that and an average acceleration profile. By observing traffic behavior from the video data at each ramp and identifying the typical distance that yellow-path vehicles cover in the three seconds upstream of the merge point, we avoid the complication of dealing with individual vehicle trajectory predictions where speeds are not constant (i.e., vehicles are accelerating), without sacrificing reasonable accuracy.

We also cannot use brake lights of yellow-path vehicles as a sufficient condition for identifying when they were too close for a red-path vehicle to proceed, as not all drivers brake when they are uncomfortably close to the vehicles in front of them—some just release the

accelerator, some choose to follow at an uncomfortable distance for a short time, and some (on these ramps in particular) simply defer their desired acceleration to freeway speeds. Another, more practical problem with using brake lights to identify when vehicles were close enough to be hazards is that some of the camera angles are looking toward oncoming traffic, such that we have no view of the brake lights.

Although our distance-based criterion may seem to inappropriately penalize those rare red-path drivers who do not yield to an approaching yellow-path vehicle within the distance threshold but then compensate by accelerating very aggressively before the yellow-path vehicle can react, these instances will also be captured by our estimates for the fraction of deliberately noncompliant drivers (β'_n), and will not affect our estimates for the long-term effectiveness of any of the strategies we test. It will, therefore, not affect our ability to make comparisons of effectiveness between strategies, either. As an aside, it would have been difficult in practice to evaluate which drivers had attempted such aggressive acceleration after improperly proceeding through the merge point, due to the limitations of some camera angles.

As mentioned previously, we do observe infrequent collisions between vehicles in the video data. When such events occur, we stop processing the video data and resume only when the incident has been resolved and the vehicles involved have left the adjacent roadway shoulder, since we suspect that their presence could have an effect on the yielding behavior of any other red-path drivers arriving during that time, which we would be unable to control for. These collisions captured in our videos were all non-injury incidents, and typically took between 15 and 25 minutes to clear.

On occasion, the freeway would suddenly become very slow near a particular ramp, possibly due to an incident that had occurred farther downstream, or a temporary traffic break created by CHP. When these situations occurred, we would cease processing the video data at the moment that one of the freeway mainline lanes was first observed to come to a (momentary) stop, and would resume only when traffic in all lanes had returned to normal free-flow conditions (typically taking up to 30 minutes). We could not process data for these periods because the distance-based thresholds were no longer appropriate when the freeway slowed down significantly (i.e., to stop-and-go speeds), since yellow-path vehicles in these circumstances often did not accelerate to the same degree that they otherwise would have if the freeway had been moving at free-flow speeds. Because the accuracy and precision of our compliance rate estimates are paramount at our experimental ramp sites, we could not afford to ignore such long spans of time in our data for these ramps. Therefore, for those weeks when such congestion occurred near an experimental ramp, we ceased data analysis entirely and used a different week at approximately the same part of the post-change analysis instead (whenever available).

For example, as we were processing the video data for June 19, 2011 (week 24 for the Yield Line “after” period analysis), we discovered two hours and 15 minutes into the video file that the freeway had become heavily congested near the experimental ramp. Consequently, we aborted our analysis of the data for June 19 and started over using the data for the following Sunday (June 26, 2011, or week 25 for the Yield Line “after” period analysis) instead. In the case of Dec 12, 2010, however, congestion occurred at the subject ramp for 30 minutes, but since there were no other feasible alternative dates to choose from, we simply disregarded the data from this 30-minute window of congestion and used the remaining 150 minutes for our analysis.

For control ramps, if the duration of congestion was found to exceed half of the 3-hour analysis period in a given week, that week of data was also disregarded. Fortunately, such occurrences of heavy, unexpected congestion tended to occur on the eastbound side of the

freeway rather than the westbound side (which was the side with most of the control ramps and all of the experimental ramps).

Considerations for Specific Cases

We counted all yields as proper yields as long as they were performed prior to entering the merge point. If a red-path vehicle came to a stop only after partially obstructing the path of the cross-traffic, it was not considered a proper yield; instead, it was encoded as a failure to yield. This situation presumably represented a driver who had not anticipated having to yield (i.e., was not prepared for it beforehand by observing signage or other controls), and was more likely responding to an imminent collision with conflicting cross-traffic rather than performing a planned or anticipated yield. This would be analogous to the situation where a driver does not notice a red light, but stops abruptly mid-intersection to avoid a possible collision with traffic on the cross-street that the driver had just noticed.

Conversely, if a red-path vehicle stopped very abruptly but prior to reaching the merge point, we had to consider that a proper yield even though it was uncertain whether the driver was aware of the Yield requirement or was simply responding to a perceived collision risk with conflicting traffic. Because some drivers naturally stop more abruptly than others, we had no way of ruling out that the driver knew to yield and just happened to be a “sudden stopper.” Also, from a practical standpoint, it would have been difficult to establish criteria for how suddenly a stop must be before it is not considered a proper yield, and would have been even more challenging to consistently apply this maximum deceleration threshold efficiently and accurately from one red-path vehicle to the next.

If a red-path vehicle arrived at the Yield point in the presence of yellow-path traffic at the merge point, we would not label it a proper yield because the red-path vehicle had no opportunity to demonstrate (through its behavior) whether it knew to yield or not—rather, the decision to yield was forced upon it, as the path before it was obstructed by conflicting traffic. However, if the red-path vehicle could have beaten any approaching yellow-path vehicles to the merge point (specifically, the red-path driver could have gotten his vehicle’s front end through the merge area before the yellow-path driver could have) and instead slowed down or stopped to let the yellow-path vehicle proceed, this would be considered a proper yield, as the red-path vehicle had the ability to proceed (inappropriately) and demonstrated its awareness of the Yield requirement by waiting instead.

Similarly, if a red-path vehicle arrived at the yield point in the presence of cross-traffic (such that the red-path vehicle stopped, but was not counted as a “proper yield”), it could subsequently have been classified as a properly yielding vehicle if it had the option to force its way into the yellow-path cross traffic via a small gap that was still large enough to squeeze into without risking a collision (i.e., a gap of about 1-2 seconds). If the red-path vehicle let such opportunities pass and instead waited for a larger gap wide enough to accommodate it without disrupting the cross-traffic flow, we considered the behavior an example of proper yielding, as the red-path vehicle made an observable choice that reflected whether the driver was aware of the Yield requirement (and corresponding right-of-way assignment) or not. Of course, if the red-path vehicle took advantage of the small gap, it was considered a failure to yield.

If a red-path vehicle arrived at the Yield point and no cross-traffic was present on the yellow path, we could not classify the driver as a proper yielder or a violator because there was no one to yield to in the first place. Even if the red-path vehicle stopped at the Yield point and waited a couple seconds to check for cross-traffic, we did not count this person as someone who

yielded properly (even though from the behavior, it was clear the person was aware of the Yield requirement) because there was no way of knowing how many of those red-path drivers who arrived at the Yield point in the absence of cross-traffic and did not stop or slow down were ignorant to the Yield requirement, and how many knew to Yield, but observed in advance that there was no approaching cross traffic and therefore did not stop or slow down. Since we can only measure proper yields and not violations when there is no yellow-path traffic present, we must ignore all red-path drivers who arrive at the Yield point under such conditions, or we would end up inappropriately inflating our estimate of the fraction of drivers that were yielding properly.

If a red-path vehicle stopped and yielded to several yellow-path vehicles or passed up multiple gaps that were large enough to squeeze into, this counted as just one instance of proper yielding. Although the driver did in fact yield to several vehicles, this is still just one “Proper Yielding” data point, since it indicates that only one driver knew what to do (as opposed to multiple drivers knowing that they were supposed to yield properly). The only situation where a driver could have been counted twice would be if the driver initially yielded properly, and after waiting a while for approaching yellow-path traffic to clear, decided to cut in (presumably out of impatience).

Data Analysis

Between 100 and 200 data points were obtained on average from each location each week, with each data point representing a red-path driver’s choice to yield or not, given conditions that warranted yielding but presented the opportunity to make either choice (i.e., to be compliant and yield properly, or to be noncompliant and proceed without yielding). Because of the large number of data points we collect each week, we are able to keep our data disaggregated so that we can more precisely understand what happens during the acclimation or novelty periods.

To assist with the raw data analysis, we developed an automated video processing system using a combination of Adobe After Effects (for video stabilization, background removal, and other pre-processing), Adobe Photoshop (for extraction of vertical and horizontal “slices” from each frame at critical coordinates), and MATLAB (for applying appropriate logic to identify when vehicles yielded properly and when they failed to yield). From start to finish, the workflow was able to fully process three hours of video in about six hours, but validation tests showed that it was not as accurate as manually processing the video footage. Furthermore, it could not be adapted to process the footage at Locations 2, 5, and 8 due to the particular camera angles for those locations. For these reasons, the automated video processing approach was abandoned in favor of manually processing each video file by watching the raw footage and recording all instances of proper yielding and failures to yield in a spreadsheet (complete with timestamps for each event).

A separate set of tools was therefore developed to assist with the manual data analysis. These tools incorporated a video viewer that could be controlled with keyboard shortcuts (including speeding up or slowing down the video in 25% increments, and jumping back or forward five seconds at a time), and a dashboard with buttons that would allow the reviewer to indicate when a proper yield or failure to yield had occurred. At the end of a session, the recorded events and their associated timestamps were saved to a text file. Windows and Mac OS X versions of the program were developed, and have been documented in greater detail in Appendices D and E. Figure 10 shows the Windows version of our video-processing tool.

Slightly modified versions of the Windows and Mac OS X viewer tool packages enabled us to record traffic volumes and speeds in place of yielding events. Another modified version

Playback controls.

*Overlay controls for Locations 5 and 8,
and synchronization tool for Location 9.*

FIGURE 10: Video processing tool developed to keep track of yielding events in each clip.

could be used to synchronize the videos at Location 9 following the procedure described in the section titled “Camera Angles and Details.” The Windows version could also be used to overlay markers on the video to indicate the appropriate distance-based thresholds for the red-path and yellow-path vehicles, using other stationary objects in the image frame (e.g., the edge of the sound wall) as registration markers. Figure 11 provides an overview of the process used for calibrating these markers and overlaying them on a video file at Location 5, and Figure 12 provides the same instructions for Location 8 footage. These were the only two locations where such overlays were needed.

Processing the Video Data

Because of the sheer volume of video footage that needed to be processed (see Figures 4-6), several undergraduate research assistants were hired to help with the effort in 2011 and 2012. A copy of one of the announcements used to recruit assistants is included in Appendix F. We met with the candidates individually before hiring them, to confirm that they were committed to the accuracy and reliability of their results rather than just looking for money. At every step of the work period, we emphasized to them that the highest priority was the accuracy of their work, not meeting deadlines, and gave them opportunities to decline to continue working if they ever felt that they no longer wanted to continue (or were no longer able to).

Next, we provided each assistant with a detailed set of instructions that were specifically written for the camera angle that person would be working on. We also provided approximately 15 sample clips showing various kinds of proper yields and failures to yield for the particular camera angle they would be watching, along with detailed notes explaining why each was coded as a proper yield, failure to yield, or non-event.

When an assistant had finished going through the instructions and sample clips, he or she was then given three hours of footage that we had already processed and an accompanying summary file of all the events contained in that clip, to use as training and calibration material. The assistant then watched the first 45 minutes of pre-processed calibration footage and followed along on the event summary sheet, noting all of the things that counted (and did not count) as proper yields and failures to yield. The last step in the training process had the assistant attempting to process the next 45 minutes of calibration footage without referencing the event summary sheet that we provided, after which the assistant was to compare his or her results to ours, review any discrepancies, and ask us for clarifications whenever necessary regarding a particular event. The assistant would then process another 45 minutes of calibration footage, compare the results to ours, review the discrepancies, ask for clarifications as needed, and repeat this process until the difference between his or her results and ours was less than 5%. Once that accuracy tolerance was met, the assistant was allowed to start working on unprocessed footage instead.

Each assistant periodically had his or her results validated through random spot-checks to confirm the continual reliability of that person’s work. For these checks, we would re-watch the entire 3 hours of footage that the person had processed, produce our own list of yielding events for that clip, and compare these results to the assistant’s. If any serious discrepancies were noted, the results from that assistant were deemed unreliable and were completely discarded (fortunately, this only happened once). Each assistant was subjected to at least two of these random accuracy checks, with more being conducted on those individuals who were responsible for processing greater amounts of footage.

FIGURE 11: Superimposed guides for the yellow-path and red-path thresholds at Location 8.

Footage showing the guides actively superimposed on the footage for Location 5.

Correspondence between the superimposed guides and the distance-based thresholds.

FIGURE 12: Interpretation of the superimposed guides at Location 5.

The results from these spot checks were then used to perform one final post-calibration on each assistant's work before incorporating his or her results into our analysis. For this, we used the percent difference between the number of yields we counted and the number of yields the assistant counted in the validation footage, and used this factor to rescale that assistant's results for all the files he or she processed. This process was repeated for failures to yield as well, with overall yield rates for each location and week calculated based on that person's rescaled number of recorded proper yields and failures to yield. These rescaling factors were within the range of 0.75 and 1.33 over 93% of the time.

5. APPLYING THE ANALYSIS FRAMEWORK

Having established a consistent and clearly defined system for evaluating driver compliance at the various on-ramp locations from week to week, we now turn to our underlying goal of evaluating the effects of different compliance strategies on the yielding behavior of drivers at these locations. The first step involves describing the existing conditions at each site and developing prediction models for each one using covariate data. We then make the experimental changes (i.e., implement the compliance strategies), monitor driver compliance rates in the weeks that follow, and use this data to evaluate the effects that the strategies have had.

Initial Survey

From a preliminary ramp survey conducted in March 2008, we found that between 50 and 60 percent of the red-path drivers failed to yield properly when yellow-path traffic was close enough to be a hazard. A second, more detailed study of existing conditions on these ramps between October 2009 and December 2009 revealed a noncompliance rate of between 38% and 51% depending on location, as shown in Figure 3.

This data for the "before" period also suggest that although the high violation rates may be partially an enforcement issue, a larger contributor to the problem is simply that drivers do not realize they are required to yield. This is supported by the observation that drivers frequently failed to yield even when doing so resulted in a high collision risk with the cross-traffic (i.e., the red-path traffic risked being rear-ended by close, fast-approaching yellow-path vehicles), while many others ended up stopping very suddenly at the merge point when cross-traffic appeared in front of them, as if it were unanticipated (i.e., a red-path vehicle stops quickly to avoid sideswiping an unexpected yellow-path vehicle arriving first to the merge point). As another indicator of the severity of the problem, collisions were also observed in the video data for Atlantic Blvd at a rate of about 1.8 per 24 hours (and at lower rates for other interchanges).

Experimental Changes to the Ramps

In November 2008, we approached the California Department of Transportation (Caltrans) with a description of the compliance problem on the ramps shown in Figure 3, and proposed several solutions to be evaluated over the next few years. Caltrans was very receptive to the idea and was eager to see the research conducted. In fact, at their request, the duration of data collection for the "after" period was extended to ensure that we would not mischaracterize the long-term effects of these compliance strategies (as some previous literature has done). After several meetings, we agreed upon a set of three compliance strategies, a set of locations, and a timeline for their implementation. As the initial proposal indicates, some of the strategies we had hoped to test were later determined to be too novel or severe for use on the ramps we were studying (i.e., a

flashing beacon or LED highlighting around the Yield signs), but they were replaced with other appropriate alternatives. Specifically, the third strategy was changed from LED highlighting to supplemental pavement text. The initial proposal is included in the appendix (see Appendix A), along with a memo from December 2010 that recommended Locations 1 and 5 for the final two compliance strategies and provided justifications for these recommendations (see Appendix B).

The first of the three compliance strategies to be implemented was larger signage at one of the ramp merges (Location 8 in Figure 3) during the week of Dec 21, 2009, increasing the width of the sign from 36 inches to 48 inches, which increased its viewable area by 78%. In the process, the sign was not repositioned, and no other elements of the experimental ramp or any of its control ramp counterparts were changed. For our analysis of the larger sign's effectiveness, Location 8 is therefore our "experimental" site, whereas Locations 1, 5, and 9 are "control" sites. Figure 13 shows the new, larger sign after being installed at Location 8.

Next, during the week of Jan 8, 2011, a Yield Line (a type of symbolic pavement markings) was installed at Location 5, for which Locations 1, 6, and 8 could function as control sites. However, to avoid potential complications associated with using a former experimental site as a current control site, Location 8 was excluded from the pool of potential control sites for the analysis of the Yield Line. Figure 13 shows Location 5 after the Yield Line was installed.

Finally, in early August 2012, Caltrans implemented the last of three planned compliance strategies: the word "Yield" written on the pavement at Location 1. Figure 13 shows this location after the pavement lettering was added. Unfortunately, this was almost a year later than the originally projected date by Caltrans, and was too late to incorporate into our analysis here. However, some "After" video data was still collected at the experimental ramp and at Location 2 (a potential control ramp) for the weeks following the installation of the pavement text, so that it would be available for analysis at a future date (see Figure 6).

Control Ramps and Other Covariates

One of the distinguishing features of this study was to use control ramps and other predictor variables to project the expected baseline compliance at the experimental locations for the post-change weeks of each strategy. Having control ramps on the same side of the freeway (e.g., Locations 1 and 5 for the Larger Sign analysis at Location 8) allowed us to capture the potential effects of things such as the angle of sunlight, which could make it harder for red-path traffic to check for approaching yellow-path vehicles, and could also make the Yield Sign and other roadway features more difficult to notice. Having a control ramp very close to the experimental ramp (e.g., Location 9 for the Larger Sign analysis) allowed us to capture the potential effects of localized environmental conditions such as construction activity (in some weeks, work was done in the freeway median near some interchanges but not others), minor differences in geometry from one interchange to another (although these interchanges share very similar designs, some have slightly tighter corners, narrower shoulders, or different signage leading up to the Yield point), and special events in the community (e.g., a large event at a nearby church or school).

In addition to the compliance rates at the experimental site and two to three control sites, we also measured several other variables for potential use in our model:

- On-ramp (entrance ramp) traffic volume.
- Natural log of on-ramp volume.
- Percent of time on-ramp flow exceeds 60, 70, or 80 vehicles in 5 minutes (5-minute bins).
- Freeway speed passing the on-ramp.
- Percent of time the freeway speed is below 55 mph (5-minute bins).

*Larger Yield sign
installed at Location 8 in December 2009.*

*Symbolic pavement markings (Yield Line)
added to Location 5 in January 2011.*

*"Yield" pavement lettering
added to Location 1 in August 2012.*

FIGURE 13: Photos of the compliance strategies implemented at Locations 1, 5, and 8.

- Percent of time direct sunlight is present (5-minute bins).

Volumes on the on-ramp were included as candidate explanatory variables with the reasoning that high volumes may result in longer wait times for red-path vehicles to find a suitable gap, leading to more inappropriate cutting behavior. Or, alternatively, higher volumes may increase the visibility of yellow-path cross traffic to red-path vehicles as they approach the yield point, making it more likely that they will realize the importance of yielding there. Since these effects might become significant only during periods of high on-ramp flow, we also had related variables to capture the percent of time the on-ramp volumes exceeded 720, 840, and 960 vph—all of which were considered possible thresholds for high on-ramp flows, especially when considering that the average volumes on the experimental ramps is typically 317 vph at Location 5 (where the Yield Line was installed) and 608 vph at Location 8 (where the larger Yield sign was installed). We also used the natural logarithm of on-ramp volume as a possible covariate, with the rationale that the marginal effect of increases in volume diminishes as they get higher. For example, the difference from 150 to 300 on-ramp vehicles may have a more powerful effect on driver behavior than an increase from 450 to 600. In some cases, counts were easily obtained from inductive loops on the on-ramps, but in other cases they had to be manually counted in the video footage. A modified version of the Video Processing Tool was used to aid with the traffic counts for those weeks that the detector data was unavailable (see Appendix G for details).

Two variables related to freeway speed were also tested, as the alignment of the red-path ramp gives those drivers a clear view of the freeway as they approach the Yield point. Consequently, unusual or unexpected traffic conditions on the mainline (such as lower speeds or higher densities) could serve as a distraction to drivers on the red-path—due to curiosity or anxiety, for instance—thereby affecting their ability to notice traffic controls (i.e., the Yield sign) on the ramps. The percent of time the freeway speed dropped below 55 mph was included with the reasoning that the distracting effect of congestion may depend less on the degree of congestion and more on the simple presence of it, given that the freeway is typically flowing freely during the analysis period. In some cases, freeway speeds were easily obtained from inductive loops on the mainline, but in other cases they had to be manually extracted from the video footage using 180 samples per week and measurements of the roadway geometry.

The sunlight variable was intended to capture any potential effects of dim (i.e., cloudy conditions) or bright light on red-path drivers' awareness of their surroundings as they approach the yield point. To measure this variable, we sampled the video footage for the ramp of interest every five minutes, using the presence or absence of shadows to determine whether the sun was out.

Variables for prevailing freeway flow and density were not included in our set of potential predictor variables, as many of the mainline inductive loops were not functioning properly for large portions of our study period, and estimating these properties from our video footage proved too difficult to be feasible.

In searching for the best possible combination of predictor variables for our $P(t)$ model (i.e., the one that minimized the prediction risk according to our cross validation), we tested every possible combination of variables (up to four at a time) rather than using a stepwise approach or similar method, since those have no guarantee of optimality. We also ensured that no more than one variable from each category below was allowed to enter the model at a time:

- Ramp volume variables.
- Freeway speed variables.

- Percent of time direct sunlight is present (this category has only one variable)
- Compliance rate at a control ramp on the same side of the freeway.
- Compliance rate at a control ramp on the opposite side of the freeway, at the same freeway exit (this category has only one variable).

For example, if a candidate model formulation had a variable for freeway speed, it could not also have a variable for the percent of time that freeway speeds dropped below 55 mph.

Ramp User Pools

Note that we do not consider differences in driver norms and habits from one site to another, which we might expect if we compared drivers from different regions or parts of town. Our study corridor is only 2.5 miles long, and is through a relatively homogenous neighborhood. Since we are collecting data on Sundays, we can assume most of the observed traffic is local, such that there are no significant differences in the user demographics from one site to another that might otherwise affect the observed effectiveness of a compliance strategy. In other words, if all covariates were held constant, we assume that if *Strategy A* performed 20% better than *Strategy B* at one entrance ramp, we would obtain a similar result if a different location on the corridor had been used instead.

Figure 14 shows the spatial distribution of several key demographic variables in the census block groups surrounding our corridor that are most likely to be served by the ramps we are studying. Note that the distance between Interstate 10 and the top of each map is approximately half the distance to Interstate 210, the next major east/west freeway in the region, such that the census tracts off the map to the north are less likely to be users of Interstate 10. Furthermore, trips generated in the lower third of each map are more likely to be captured by State Route 60 near the bottom, rather than Interstate 10. Therefore, the population of interest is predominantly confined to the upper two thirds of each map.

Of biggest concern is whether the demographics of the area change significantly from left to right, as the user pools for each ramp depend on the census blocks that are closest to each arterial. For example, if the census blocks surrounding Atlantic Bl have significantly different demographics from those surrounding San Gabriel Bl, it implies differences in the drivers using the ramps at each of those two locations.

It is less critical that the population be consistent on the southern side of the freeway, or between the two sides of the freeway, because of the way we have designed this study. Drivers starting their trips to the south of the freeway would only be captured by our analysis if they entered the freeway heading east (because the ramps connecting the northbound arterials to the westbound side of the freeway do not include yield signs), whereas drivers starting their trips to the north of the freeway would only be captured if they entered the freeway heading west. Since all of our compliance measures were implemented on the westbound side of the freeway, the resultant MOEs that we compute depend only on the behavior of the drivers coming from the census tracts north of the freeway.

The maps in Figure 14 indicate how the following four key demographic characteristics change across the capture zones for the different arterials along the Interstate 10 corridor:

- Education, quantified by the percent of people 25 years or older who graduated high school.
- Area Desirability, quantified by the median value of owner-occupied homes.
- Ethnic Background, quantified by the percent of the population that is foreign-born.
- Income, quantified by the percent of the population below the poverty line.

Percent of persons 25 years and older with a High School Diploma or More Education.

Percent of families below the poverty level.

Percent of persons who are foreign born.

Median value of specified owner-occupied housing units.

FIGURE 14: Demographic information for the area surrounding the study corridor, from the 2000 census.

For the area to the north of the freeway, the population characteristics do not change appreciably from left to right except with respect to education, where there is a slight trend of increasing high school graduation rates to the west. However, in the vicinity of Locations 5 and 8, which we are currently using to analyze the effects of larger signage and a Yield Line, the population is relatively consistent across all four demographic variables.

Prediction Model Formulation

To evaluate the predictive power of each model formulation, we will apply Leave-One-Out Cross Validation (LOOCV), which is a specialized type of K-Fold Cross Validation. In LOOCV, one data point (i.e., one week's compliance rate) is used for validation and all remaining data points are used for training. The process is repeated so that each data point is excluded from the training set once, and the resultant risk from each iteration is used to compute an average risk for the overall model. When we have 10 bins of data, the LOOCV procedure is identical to 10-Fold Cross Validation.

For our risk estimates, we will use a squared error loss function. We estimate the uncertainty associated with our predictions (σ_{cv}) using the standard deviation of the residuals between the predicted and observed values in each LOOCV iteration. We also evaluate the Durbin Watson statistic to test for serial correlation in our time series data, which we can then properly address if necessary. Finally, we compare the predictive capability of our risk-minimizing model to a constant-only model using a paired one-tailed *t*-test.

One simplification we make in applying this procedure is that we use an arithmetic mean across all LOOCV iterations, rather than a weighted average for the risk associated with each iteration based on the number of yield events (i.e., proper yields and failures to yield) observed in each week. Consequently, those weeks where the observed yield rate is more precise (due to being calculated from a greater number of observed yield events) are not given more weight than the less-precise compliance rates evaluated for those weeks where fewer yield events were recorded. This simplification avoids additional complications associated with evaluating σ_{cv} or with carrying out the paired *t*-tests.

Fortunately, this simplification can be justified by noting that each of the optimal prediction models identified in this dissertation (specifically, Equations 10, 11, 12, and 15, which will all be shown later) did not change when our analyses were repeated using weighted averages in place of simple arithmetic means for evaluating the prediction risks associated with the LOOCV procedure. However, the same may not hold true for other compliance strategy evaluations; therefore, when applying this framework, one should generally use weighted averages for evaluating the prediction risk when possible, particularly if compliance rates are based on a number of events that varies significantly from one bin to another. In the case of our analyses, the number of events recorded was fairly consistent from week to week. For the weekly Larger Sign “before” data, the number of events each week averaged 168, with a standard deviation of 14.9 (or 8.9% of the mean). For the weekly Ramp Rehabilitation and Yield Line “before” data, the number of events averaged 103, with a standard deviation of about 15.0 (or 14.6% of the mean).

Driver Response Curve

In evaluating the driver response curve, we tested two possible distributions for the Awareness Effect: a Normal distribution and a Log-Normal distribution. The reasoning behind a Normal distribution was that the amount of time that it takes a driver to notice something new in

his or her environment is an inherent characteristic of the driver, and that it could be expected to follow a Normal distribution just as other characteristics of a person would (e.g., height). The reasoning behind a Log-Normal distribution was that the time it takes someone to notice a change in their driving environment may follow the same distribution as the reaction time for things that require action (such as a traffic signal turning red), which is generally treated in literature as following a Log-Normal distribution.

For the Novelty Persistence curve, we tested three possible distributions: a Normal distribution, a Geometric distribution, and a Uniform distribution. The reasoning behind a Normal distribution was based on the same logic given earlier for the Awareness Effect distribution. We also reasoned that drivers might have a certain, constant probability of no longer finding the strategy novel each time they are exposed to it, such that the persistence of its novelty would follow a Geometric distribution (i.e., the number of trials until “success,” where success represents a driver that no longer thinks of the change as something novel). Finally, we tested a Uniform distribution with the reasoning that equal numbers of drivers may stop considering the strategy to be novel each week, until there are no more drivers left that exhibit a Novelty Effect. We tested all combinations of Awareness Effect and Novelty Persistence distributions, and selected the one that resulted in the highest value for the likelihood function.

When specifying a Normal distribution for the Awareness Effect or Novelty Persistence, we solve for the proportionality constant (β_3) as one of our shape parameters, where $\sigma_a = \beta_3 \mu_a$ for $f_a(t_a)$ when Normally distributed, and $\sigma_n = \beta_3 \mu_n$ for $f_n(t_n)$ when Normally distributed. When specifying a Geometric distribution for Novelty Persistence, we used the relationship $p = (\mu_n)^{-1}$. Finally, when specifying a Log-Normal distribution for the Awareness Effect, we converted μ_a and σ_a from the Normal distribution parameters into α and β Log-Normal shape parameters according to the following relationships:

$$\beta = \sqrt{\ln\left(\frac{\sigma_a^2}{\mu_a^2} + 1\right)} \quad (8)$$

$$\alpha = \ln(\mu_a) - \frac{1}{2}\beta^2 \quad (9)$$

Because of the highly non-linear nature of the Driver Response Curve equation, common optimization tools such as the built-in Excel Solver were ineffective at finding the optimal parameter values. Thus, we used a straightforward brute-force optimization approach that discretized the domain (i.e., shape parameters) by small increments and tested each combination of parameters. The range of values tested for each parameter was limited to include only those that were deemed sensible. The following is a summary of the ranges used for each variable:

β'_a (long-term reduction in violators): 0–100%

β'_2 (fraction of deliberately noncompliant group exhibiting a novelty response): 0–100%

μ_a (mean of the Awareness distribution): 0 weeks minimum, 5–15 weeks maximum.

μ_n (mean of the Novelty Persistence distribution): Same as μ_a

β_3 (proportionality constant for Normal distributions): 0.3–1.0

The optimization occurred in two passes: the first pass broke the range of each variable into approximately 10–20 increments, and the second pass used the optimal values from the coarse optimization (i.e., first pass) and reduced the increments by roughly one order of magnitude (using a new domain that was ± 1 coarse-level increment away from the optimal values of the first pass) to find a more precise optimum.

Estimating β'_n

To estimate (β'_n), we need a situation where we can reasonably assume that all drivers were aware of the required behavior, such that all violations were the result of a conscious decision to do so, rather than a consequence of driver ignorance. We also seek a situation that resembles the Yielding environment at our study sites in as many ways as possible. Therefore, to estimate the fraction of drivers who were deliberately noncompliant at each experimental ramp (i.e., the southbound arterial connection to the westbound freeway), we monitored corresponding return traffic (i.e., eastbound freeway traffic exiting north on the arterial) for the same mid-day period and watched to see which vehicles failed to yield properly at the exit ramp's connection to the arterial. Since the exit ramps terminate at T-intersections with the arterial (see Figure 2), it should have been apparent to all exiting vehicles that traffic on the through street (i.e., the arterial) had the right of way, as this is a fundamental driving rule in California. Thus, exiting drivers were considered deliberately noncompliant if they proceeded through the T-intersection and onto the arterial when approaching through traffic was less than 3 seconds away. Note that the satellite imagery used in Figure 2 is slightly outdated (though it is the most current image available through the USGS), and that at the time of the stop sign noncompliance analyses, the exit ramp intersections were true T-intersections with perpendicular approaches.

This right-turn situation resembles the situation at our study sites in that any deliberate failure to comply involves a merging vehicle forcing its way into a cross-traffic stream with the knowledge that cross-traffic will likely slow down to accommodate them as needed, rather than maintain its current speed and collide with the merging vehicle or follow it at an unsafe distance. Furthermore, both situations also label as violators those drivers who cannot properly gauge gap sizes and consistently merge into gaps that are too small to accommodate them. These people may in fact be aware of the yielding requirement and may simply be unaware that their behavior counts as violations, such that a compliance strategy designed to make the control more visible would not be capable of improving their behavior regardless of how effective it was. However, they would also be considered violators of the stop signs based on the way we are estimating (β'_n), such that their behavior would not inappropriately bring down the measured effectiveness of any compliance strategies being analyzed (see Equation 6). At our study sites, these people (if they exist) would represent a unique subgroup of (β'_n), since their non-compliance is neither a consequence of ignorance regarding the Yield requirement nor lack of enforcement, and is therefore difficult to address. We assume that such drivers are not a significant part of (β'_n), since such drivers would quickly learn (through their experience making Left Turns, executing passing maneuvers, etc.) how to more accurately gauge gap sizes. In this way, we may still treat β'_2 as having a range between 0 and 1. Without this assumption of insignificance regarding the size of this unique subgroup of β'_n that cannot accurately gauge gaps, the upper limit on β'_2 would have been some value slightly less than 1, depending on the fraction of β'_n that is comprised of this subgroup.

To follow the approach used for the on-ramp locations as much as possible, the criterion used to determine whether or not a right-turning vehicle should have proceeded will be distance-

based and will require free-flowing traffic conditions on the cross street. This ensures that those rare instances where a driver chose to enter a gap that was too small and then aggressively accelerated to avoid affecting the cross-traffic (which arguably may not count as a failure to yield), will be included in the estimate for (β'_n) , so that they will not improperly influence our measurement of the long-term effectiveness of a given strategy according to Equation 6.

Analysis of Larger Sign

Prediction Model Formulation

Using 10-fold cross-validation to test all possible combinations up to four covariates (and a constant term), we identified the optimal prediction model (i.e., the model that minimized the prediction risk) for Location 8 as:

$$P(t) = \text{Logit}^{-1}[-0.032x_1(t) - 0.206x_2(t) - 0.739x_3(t) + 2.619] \quad (10)$$

where

$x_1(t)$ is the average freeway speed for week t .

$x_2(t)$ is the percent of time it was sunny for week t .

$x_3(t)$ is the compliance rate at Location 5 in week t , where the gap for the week of Nov 15 (see Figure 4) is filled with an average of the rates from all other weeks.

The following table lists the properties of the model coefficients, with the raw data given in Appendix H.

Variable	Coefficient	Standard Error	Significance
Freeway Speed passing on-ramp	-0.0319	0.0127	0.0229
Percent of time the weather is sunny	-0.2057	0.1436	0.1010
Compliance rate at Location 5 Control Ramp	-0.7386	0.6192	0.1390
Constant	2.6191	1.0272	0.0218

This model (i.e., Equation 10) had an $R^2 = 0.615$, an average prediction risk of 0.000457, a prediction error mean of $\mu_{cv} = 0.00271$, and a prediction error standard deviation of $\sigma_{cv} = 0.02286$. For comparison, a constant-only model for Location 8 had an average risk of 0.000714, a prediction error mean of $\mu_{cv} = 0.003245$, and a prediction error standard deviation of $\sigma_{cv} = 0.02782$. From a paired t-test, the hypothesis that Equation 10 does no better than the constant-only model had a p-value = 0.124. Equation 10 also had a Durbin Watson statistic of 1.43, which did not provide evidence of serial correlation in the data.

The coefficients in Equation 10 all had p-values lower than 0.14, with the constant term and the coefficient for $x_1(t)$ being significant on a 5% level. In interpreting the coefficients, we see that compliance tends to increase when the speed of the freeway decreases, which may

indicate that drivers break out of “autopilot” and are more likely to check for cross-traffic when freeway conditions are contrary to their expectations. The coefficient for $x_2(t)$ reveals that compliance tends to increase when the weather becomes more cloudy, suggesting that glare from the sun may make it harder for drivers to check thoroughly for approaching cross-traffic (which is sensible considering that red-path drivers at Location 8 must look southeast when checking for yellow-path vehicles, and at midday the sun in Los Angeles is always to the south).

The coefficient for $x_3(t)$ indicates that the compliance rate tends to increase at the experimental ramp when it decreases at Location 5, which is somewhat unexpected and may require additional research to explain, though it could be the result of slight design differences in the two locations. For example, one possible explanation would be that the on-ramp at Location 8 is adjacent to a traffic signal, whereas the on-ramp at Location 5 is not. Consequently, there may be a greater tendency of vehicles to form platoons as they enter at Location 8, making these drivers more susceptible to improper “cutting” behavior when traffic volumes rise (due to seasonal effects) because of the psychological pressure associated with having other drivers waiting in line behind them as they look for gaps. At Location 5, however, increases in overall volume may make yellow-path traffic more visible to approaching red-path vehicles (which are more likely to be random arrivals than they were at Location 8), leading to greater awareness of the need to check for cross-traffic and yield to it. Again, though, this negative coefficient is contrary to our expectations and may warrant further research to fully interpret and explain.

Estimate for β'_n

To estimate β'_n , we used video footage to monitor exiting traffic from Interstate 10 eastbound to San Gabriel Blvd northbound during the 11 AM to 2 PM midday period. This exit ramp was chosen in an attempt to capture the same drivers as those that would have passed through Location 8 at some point (i.e., those drivers who would have entered the freeway heading west from southbound San Gabriel Blvd). If a vehicle arrived at the end of the off-ramp in the presence of cross-traffic on the arterial, we recorded whether it waited for a sufficiently large gap to turn onto the cross street (i.e., a gap of at least 3 seconds) or forced its way into a smaller gap that did not meet this minimum threshold. We used a minimum feasible gap size of 1.3 seconds (measured from the rear bumper of the leading vehicle to the front bumper of the following vehicle), meaning that any gaps smaller than that were not considered large enough for exiting traffic to forcibly cut into. This was based on the approximate minimum gap size that we observed on-ramp red-path traffic inappropriately squeezing into.

As with the on-ramp footage, our gaps were measured using distance-based thresholds that were superimposed on the video footage for easier identification. We used a turning threshold of 45° to define the precise moment that an exiting vehicle was considered to have entered the arterial traffic stream, and only considered traffic in the right lane of the arterial when monitoring for gaps. We also did not consider whether the exiting ramp vehicle ended its turn in the right lane (legal) or the left lane (illegal), because in either case, the maneuver would cut across the path of approaching cross-traffic in the right lane of the arterial and would elicit a reaction from any approaching drivers in the right lane that were within three seconds of the T-intersection. We also did not consider any situations involving approaching cross traffic that had recently turned onto the arterial from a nearby driveway or cross-street, since these vehicles tended to approach the exit ramp T-intersection at far lower speeds than typical arterial traffic, making the distance-based thresholds inadequate for our purposes in these cases.

Having sampled 210 drivers on this exit ramp who had the opportunity to choose between yielding properly to arterial traffic (compliant), or proceeding without waiting (noncompliant), we obtained an estimate of $\beta'_n = 0.110$ for the population of all drivers at this interchange on Sundays during the midday period, with an uncertainty of $\sigma_{\beta'_n} = 0.022$.

Results

Figure 15 compares the observed compliance rates and the predicted compliance rates $P(t)$ for the weeks before and after the larger signage was installed, where $P(t)$ is given by Equation 10. The error bars are 90% confidence intervals for the predicted values. The “before” weeks are included on the left side to give a sense of the inherent variability of the data and the strong agreement between the predicted and observed compliance rates, so that we may be confident that the differences observed in the “after” weeks on the right side of Figure 15 are not simply a consequence of stochastic fluctuations in the data. Because we observe greater compliance rates than the baseline values $P(t)$ in the weeks after the change, we have evidence that the larger signage had a measurable positive effect on driver behavior.

The improvement by week (per Equation 1) is shown in Figure 16, where the black dots represent the observed improvements and the error bars are 95% confidence intervals for these observed values (the error comes from the uncertainty associated with $N(t)$). The solid line is the overall Driver Response Curve (Equation 2) that has been fitted to the data using likelihood maximization, with the contribution from the transient Novelty Effect (i.e., Group 2 drivers) shown by a dashed line and the contribution from the enduring Awareness Effect (i.e., Group 1 drivers) shown by a dotted line.

Note that the curves in Figure 16 are approximate due to the fact that values for each week are dependent upon $P(t)$, which fluctuates from week to week and cannot be computed for those weeks lacking observed data for the covariates (e.g., weeks 7 and 10). For illustrative purposes, in Figure 16, $P(t)$ is held at a constant value of 44.13% for all weeks after the change, which is the average $P(t)$ value over the weeks in which prediction data were available. These approximations stay within 5% of the true Driver Response model at all times, and were applied only for Figure 16—and not for any part of the analysis. The result is a smooth Driver Response Curve that highlights improvements due to Novelty and Awareness effects, and excludes the effect of small fluctuations or gaps in the covariate data.

Model Parameters

For $f_a(t_a)$, we tested a Normal distribution (with $\sigma_a = \beta_3\mu_a$) and a Log-Normal distribution (with shape parameters according to Equations 8 and 9). For $f_n(t_n)$, we tried three different distributions: a Uniform distribution having mean μ_n , a Geometric distribution (with shape parameter p_n), and a Normal distribution (having mean μ_n and assuming $\sigma_n = \beta_3\mu_n$). We used 12 weeks as an upper limit on the mean of $f_a(t_a)$ and $f_n(t_n)$, which proved to be a non-binding constraint.

The combination that maximized the likelihood function was a Normal distribution for the Awareness Effect and a Uniform distribution for the Novelty Persistence effect. The parameters that resulted in a maximum Log-Likelihood of -15.093 were:

$$\beta'_a = 0.19$$

$$\beta'_2 = 1.00$$

FIGURE 15: Observed and predicted compliance by week at Location 8 (larger sign).

FIGURE 16: Percent reduction in violation rate by week at Location 8 (larger sign).

$$\mu_a = 3.7 \text{ weeks}$$

$$\mu_n = 5.0 \text{ weeks}$$

$$\beta_3 = 0.83$$

Table 1 shows the likelihood function values across all increments of the proportionality constant β_3 , for each combination of Awareness and Novelty Persistence distributions. The optimal parameters are shown across these three variables (i.e., the value for β_3 and the distribution families for $f_a(t_a)$ and $f_n(t_n)$), as this will aid us later when we attempt to unify our separate results from the Larger Sign, Ramp Rehabilitation, and Yield Line analyses. Interestingly, the Log-Normal distribution performed worse than the Normal distribution at modeling the Awareness Effect in every instance when $f_n(t_n)$ was assumed to follow a Uniform distribution, and all but one instance (i.e., when $\beta_3 = 0.95$) when $f_n(t_n)$ was assumed to be Normally distributed, suggesting that the Normal distribution is overall a better model for $f_a(t_a)$ than a Log-Normal distribution is.

Based on Equation 6, we found the larger signage to be 19% effective at reducing the fraction of noncompliant drivers in the long run. We will use likelihood ratio tests to establish confidence intervals for this MOE after we have obtained results for the Yield Line at Location 5.

Our optimal value for β_3 is slightly higher than the 0.5 suggested in the literature (10), but it is still reasonable. Because the value of β'_2 hit its upper limit of 1.00 during likelihood maximization, it is possible that this value would have gone higher had the constraint been relaxed. This suggests that the true value of β'_n may be higher than our estimate of 0.11.

Because the shape of the Novelty Effect curve (see Figure 16) depends on the Awareness Curve (see Equation 4), its influence can last much longer than $f_n(t_n)$ itself suggests. In the case of a larger sign, the Novelty Effect continued to play a meaningful role in the behavioral response of drivers through Week 18, even though the Novelty Persistence had a mean of only $\mu_n = 5$ weeks. We also note that the Driver Response Curve does not start off strong and then decline over time as is generally assumed in the literature. Rather, it builds up gradually to a peak before tapering off to its long-run value. This is supported by both our Driver Response Model and our observed data.

Significance of Parameters

To examine whether each shape parameter is significant, we force each parameter to zero, solve for the remaining parameters using likelihood maximization, and perform a likelihood ratio test. From this procedure, we obtain the following p-values: 0.000 for μ_a and β_3 , and 0.220 for β'_2 and μ_n . Therefore, the Driver Awareness Effect is significant on a 5% level, though our data does not show a statistically significant Novelty Effect even at a 10% level. Figure 17 shows the Driver Response Curves using the parameters that maximized the likelihood function when either μ_a and β_3 or β'_2 and μ_n were constrained to zero, to visualize the significance of their effects on the Driver Response Curve.

Note that β_3 and μ_a must be removed from the driver response model simultaneously, because β_3 becomes unnecessary without μ_a . As a consequence, the Chi-squared statistic used for this likelihood ratio test has two degrees of freedom. The same is true of β'_2 and μ_n .

$f_a(t_a) : \text{Normal}$							$f_n(t_n) : \text{Normal}$						
μ_n	β'_2	μ_a	β_3	β'_a	Log Likelihood	Rank	μ_n	β'_2	μ_a	β_3	β'_a	Log Likelihood	Rank
0.5	0.00	1.5	0.30	0.22	-17.875	14	1.5	0.05	2.0	0.30	0.22	-18.074	14
0.5	0.35	2.0	0.35	0.22	-17.455	13	1.0	0.05	2.0	0.35	0.22	-17.719	13
0.5	0.55	2.0	0.40	0.22	-17.098	12	0.5	0.20	2.0	0.40	0.22	-17.420	12
0.5	0.65	2.0	0.45	0.22	-16.829	11	0.5	0.30	2.0	0.45	0.22	-17.161	11
0.5	0.75	2.0	0.50	0.22	-16.635	10	0.5	0.35	2.0	0.50	0.22	-16.937	10
3.5	1.00	4.5	0.55	0.22	-16.574	9	0.5	0.30	2.0	0.55	0.22	-16.833	8
3.5	1.00	4.0	0.60	0.22	-16.317	8	1.0	0.15	2.0	0.60	0.22	-16.867	9
3.5	1.00	4.0	0.65	0.22	-15.929	6	1.0	0.60	2.5	0.65	0.22	-16.772	7
3.5	1.00	4.0	0.70	0.22	-15.782	5	1.0	0.55	2.5	0.70	0.22	-16.748	6
5.0	1.00	3.5	0.75	0.18	-15.745	3	1.0	1.00	3.0	0.75	0.22	-16.726	5
5.0	1.00	3.5	0.80	0.18	-15.570	2	1.0	0.95	3.0	0.80	0.22	-16.712	4
5.0	1.00	3.5	0.85	0.18	-15.555	1	12.0	1.00	3.0	0.85	0.10	-16.536	3
5.0	1.00	3.5	0.90	0.18	-15.773	4	12.0	1.00	3.0	0.90	0.10	-16.160	2
4.5	1.00	3.5	0.95	0.18	-15.997	7	12.0	1.00	3.0	0.95	0.10	-15.880	1
4.5	1.00	3.6	0.80	0.19	-15.390	Optimal	12.5	1.00	3.0	0.95	0.10	-15.690	Optimal

$f_a(t_a) : \text{Normal}$							$f_n(t_n) : \text{Uniform}$						
μ_n	β'_2	μ_a	β_3	β'_a	Log Likelihood	Rank	μ_n	β'_2	μ_a	β_3	β'_a	Log Likelihood	Rank
0.5	0.00	1.5	0.30	0.22	-17.875	14	1.0	0.05	2.0	0.30	0.22	-18.071	14
0.5	0.35	2.0	0.35	0.22	-17.455	13	1.0	0.10	2.0	0.35	0.22	-17.718	13
0.5	0.55	2.0	0.40	0.22	-17.098	12	0.5	0.20	2.0	0.40	0.22	-17.420	12
0.5	0.65	2.0	0.45	0.22	-16.829	11	0.5	0.30	2.0	0.45	0.22	-17.161	11
0.5	0.75	2.0	0.50	0.22	-16.635	9	0.5	0.35	2.0	0.50	0.22	-16.937	10
0.5	0.75	2.0	0.55	0.22	-16.677	10	0.5	0.30	2.0	0.55	0.22	-16.833	8
3.5	1.00	4.0	0.60	0.22	-16.332	8	1.0	0.15	2.0	0.60	0.22	-16.850	9
3.5	1.00	4.0	0.65	0.22	-15.917	7	1.0	0.60	2.5	0.65	0.22	-16.775	7
3.5	1.00	4.0	0.70	0.22	-15.719	6	1.0	0.55	2.5	0.70	0.22	-16.724	6
5.0	1.00	4.0	0.75	0.18	-15.608	5	1.0	1.00	3.0	0.75	0.22	-16.674	4
5.5	1.00	3.5	0.80	0.18	-15.425	2	1.0	0.95	3.0	0.80	0.22	-16.680	5
5.5	1.00	3.5	0.85	0.18	-15.388	1	5.5	1.00	4.0	0.85	0.18	-16.465	3
5.5	1.00	3.5	0.90	0.18	-15.465	3	5.5	1.00	4.0	0.90	0.18	-16.140	2
5.5	1.00	3.5	0.95	0.18	-15.608	4	5.5	1.00	4.0	0.95	0.18	-15.869	1
5.0	1.00	3.7	0.83	0.19	-15.093	Optimal	5.2	1.00	4.4	0.95	0.19	-15.673	Optimal

TABLE 1 (Part 1 of 2): Optimal parameters and associated Log-Likelihoods for each combination of β_3 , $f_a(t_a)$, and $f_n(t_n)$ at Location 8 (for the driver response curves for the Larger Sign). Higher-resolution overall optimal values are given in the last row of each table.

$f_a(t_a)$: Normal				$f_n(t_n)$: Geometric			$f_a(t_a)$: Log-Normal				$f_n(t_n)$: Geometric		
μ_n	β'_2	μ_a	β_3	β'_a	Log Likelihood	Rank	μ_n	β'_2	μ_a	β_3	β'_a	Log Likelihood	Rank
0.12	0.00	1.5	0.30	0.22	-17.875	14	0.90	0.05	2.0	0.30	0.22	-18.073	14
0.90	0.45	2.0	0.35	0.22	-17.482	13	0.90	0.15	2.0	0.35	0.22	-17.722	13
0.90	0.65	2.0	0.40	0.22	-17.160	12	0.90	0.30	2.0	0.40	0.22	-17.430	12
0.90	0.85	2.0	0.45	0.22	-16.921	9	0.90	0.40	2.0	0.45	0.22	-17.179	11
0.90	0.95	2.0	0.50	0.22	-16.747	7	0.90	0.45	2.0	0.50	0.22	-16.962	10
0.90	1.00	2.0	0.55	0.22	-16.869	8	0.90	0.45	2.0	0.55	0.22	-16.810	4
0.12	0.00	1.5	0.60	0.22	-17.005	11	0.70	0.25	2.0	0.60	0.22	-16.888	7
0.16	1.00	3.5	0.65	0.18	-16.937	10	0.80	1.00	2.5	0.65	0.22	-16.860	6
0.16	1.00	3.5	0.70	0.18	-16.530	5	0.80	1.00	2.5	0.70	0.22	-16.828	5
0.16	1.00	3.5	0.75	0.18	-16.281	3	0.80	1.00	2.5	0.75	0.22	-16.896	8
0.16	1.00	3.5	0.80	0.18	-16.158	2	0.60	1.00	3.0	0.80	0.22	-16.912	9
0.16	1.00	3.5	0.85	0.18	-16.133	1	0.12	1.00	3.5	0.85	0.14	-16.676	3
0.16	1.00	3.5	0.90	0.18	-16.298	4	0.12	1.00	3.5	0.90	0.14	-16.431	2
0.16	1.00	3.5	0.95	0.18	-16.533	6	0.12	1.00	3.5	0.95	0.14	-16.224	1
0.17	1.00	3.4	0.84	0.18	-16.099	Optimal	0.10	1.00	3.6	1.00	0.14	-15.447	Optimal

TABLE 1 (Part 2 of 2): Optimal parameters and associated Log-Likelihoods for each combination of β_3 , $f_a(t_a)$, and $f_n(t_n)$ at Location 8 (for the driver response curves for the Larger Sign). Higher-resolution overall optimal values are given in the last row of each table.

FIGURE 17: Re-optimized Driver Response Curves at Location 8 (larger sign) when excluding either (a) the Novelty Effect or (b) the Driver Awareness Effect.

Analysis of Ramp Rehabilitation

As mentioned in the “Data Overview” section earlier, each of the entrance ramps along the I-10 corridor was rehabilitated at some point in late 2010 or early 2011. The rehabilitations involved reconstructing the pavement, pouring new curbs and gutters, and general landscaping, but did not include any changes to the layout and design of the ramps with the exception of wider shoulders and improved lighting at night. Figure 18 shows a “before” and “after” pair of images for the rehabilitation work at Location 6. These rehabilitations resulted in the closure of each ramp for 4-8 weeks as shown in Figures 4-5.

We expected that the driver response would be limited to a Novelty Effect only, given the superficial nature of the changes at each rehabilitated ramp. However, the video data revealed that some red-path drivers gradually learned that they could use the wider shoulder as a sort of temporary acceleration lane to give them time and space to negotiate merging with yellow-path traffic (when present) instead of properly waiting for an appropriate gap at the yield point. Drivers who chose to do this were considered in violation of the yielding requirement, and contributed to slightly lower compliance rates at the entrance ramp after the ramp rehabilitation had occurred.

We will be examining the effects of ramp rehabilitation at Location 5, so that we can reliably control for the effect of a wider shoulder when we do our analysis of the Yield Line at this location later. This rehabilitation occurred between the weeks of July 19 and August 20, 2010. We also assume that any reduction in compliance at this location resulting from the ramp rehabilitation could also be expected at other locations following their rehabilitation, given that the same changes occurred in each case (e.g., pouring of new pavement, landscaping, widening of shoulders).

Prediction Model Formulation

Using LOOCV to test all possible combinations of up to four covariates as we did in the larger sign analysis, we identified the following optimal prediction model for Location 5:

$$P(t) = \text{Logit}^{-1}[-1.059x_1(t) + 0.072x_2(t) - 1.605x_3(t) + 8.433] \quad (11)$$

where

$x_1(t)$ is the natural logarithm of the on-ramp volume in week t .

$x_2(t)$ is the compliance rate at Location 1 in week t .

$x_3(t)$ is the compliance rate at Location 6 in week t .

The following table lists the properties of the model coefficients for Equation 11, with the raw data given in Appendix H.

Variable	Coefficient	Standard Error	Significance
Natural Log of On-Ramp Flow for the full 3 hour period	-1.059	0.200	0.002
Compliance rate at Location 1 Control Ramp	0.072	0.548	0.450
Compliance rate at Location 6 Control Ramp	-1.605	0.855	0.060
Constant	8.433	1.204	0.000

*Sept 26, 2010
(last week before closure for rehabilitation).*

*Nov 14, 2010
(first week after reopening).*

FIGURE 18: Condition of Location 6 before and after rehabilitation occurred in 2010.

This model (i.e., Equation 11) was based on nine weeks of data rather than 10, as we had no data at Location 5 during the week of Nov 15. We also did not make Location 8 available as a control site, to avoid potential complications due to the installation of a larger sign at the end of 2009.

Equation 11 had an $R^2 = 0.923$, an average prediction risk of 0.000447, a prediction error mean of $\mu_{cv} = -0.0006$, and a prediction error standard deviation of $\sigma_{cv} = 0.02242$. For comparison, a constant-only model for Location 5 had an average risk of 0.002834, a prediction error mean of $\mu_{cv} = 0.0000$, and a prediction error standard deviation of $\sigma_{cv} = 0.05646$. From a paired t -test, the hypothesis that Equation 11 does no better than the constant-only model had a p -value = 0.032, giving us confidence in this model. Equation 11 also had a Durbin Watson statistic of 2.27, which did not provide evidence of serial correlation in the data.

The p -values for the coefficients in Equation 11 were 0.0016 for $x_1(t)$, 0.4501 for $x_2(t)$, 0.0597 for $x_3(t)$, and 0.000 for the constant. Thus, although the inclusion of Location 1 as a control site contributed to better predictions, the variable did not prove to be significant at a 10% level, whereas all other covariates were. We will discuss what happens when this variable is removed from Equation 11 in the following section, titled “Alternate Prediction Model.”

In interpreting the coefficients, we see that compliance tends to decrease when the on-ramp volume increases. This may be a result of red-path drivers becoming more aggressive with respect to gap acceptance as yellow-path volume increases and larger gaps become less frequent. The coefficient for $x_2(t)$ is positive as we expected. This implies that other factors we have not explicitly accounted for with our covariates tend to influence driver behavior in similar ways at Location 1 and the experimental ramp. Weeks where compliance dropped at Location 1 were weakly associated with a drop in compliance at the experimental site, and vice versa.

As was the case in Equation 10, the coefficient for $x_3(t)$ indicates that the compliance rate tends to increase at the experimental ramp when it decreases at Location 6. This suggests that there may be some unidentified environmental influence that affects drivers at Location 5 and Location 6 differently. For example, it could be that strong winds blowing from the west (which happens on occasion) help eastbound yellow-path traffic accelerate faster at Location 6, but make it difficult for westbound yellow-path traffic to accelerate at Location 5. The result would be gaps that appear larger than they are (in terms of headway) for red-path vehicles at Location 6, but the reverse at Location 5. Again, however, this negative coefficient is contrary to our expectations, and requires additional research for proper interpretation.

Alternate Prediction Model

Because Location 1 became unavailable in the later portion of the “after” period for the ramp rehabilitation analysis (see Figure 4-5), we also solve for an alternate model to Equation 11 that does not include Location 1 as a control ramp, for use in those weeks when Location 1 was unavailable. Using LOOCV as before, we identify our new risk-minimizing prediction model as:

$$P(t) = \text{Logit}^{-1}[-0.991x_1(t) - 0.845x_2(t) - 1.307x_3(t) + 7.803] \quad (12)$$

where

$x_1(t)$ is the natural logarithm of the on-ramp volume in week t .

$x_2(t)$ is the percent of time the freeway speed is below 55 mph in week t .

$x_3(t)$ is the compliance rate at Location 6 in week t .

The following table lists the properties of the model coefficients for Equation 12, with the raw data given in Appendix H.

Variable	Coefficient	Standard Error	Significance
Natural Log of On-Ramp Flow for the full 3 hour period	-0.991	0.155	0.001
Percent of time the freeway speed is below 55 mph	-0.845	0.534	0.087
Compliance rate at Location 6 Control Ramp	-1.307	0.689	0.058
Constant	7.803	0.979	0.000

Equation 12 had an $R^2 = 0.948$, an average prediction risk of 0.0006261, a prediction error mean of $\mu_{cv} = -0.0026$, and a prediction error standard deviation of $\sigma_{cv} = 0.02640$. The risk measurements for the constant-only model were reported in the previous section. From a paired t -test, the hypothesis that Equation 12 does no better than the constant-only model had a p -value = 0.061, which is higher than that for Equation 11, but still low enough to justify choosing it over the constant model. Equation 12 also had a Durbin Watson statistic of 1.69, which did not provide evidence of serial correlation in the data.

The p -values for the coefficients in Equation 12 were 0.0007 for $x_1(t)$, 0.0873 for $x_2(t)$, 0.0580 for $x_3(t)$, and 0.000 for the constant, such that all parameters were significant at a 10% level. The coefficients for $x_1(t)$ and $x_3(t)$ have the same signs as they did in Equation 11 as expected, and the coefficient for $x_2(t)$ is negative, meaning that compliance tends to decrease when congestion increases. This is somewhat contrary to what we saw in Equation 10, and may be related to the negative correlation between compliance at Location 5 and Location 8 as revealed by the coefficient of $x_3(t)$ in the prediction model for the Larger Sign (i.e. Equation 10). One possible explanation for this would be that yellow-path traffic accelerates differently between Locations 5 and 8, due to differences in the lengths of the acceleration zone for the yellow path (see Figure 2). Specifically, the auxiliary lane is slightly longer at Location 8. Consequently, when the freeway becomes congested, yellow-path traffic may be able to accelerate to its desired cruising speed before crossing the red-path distance threshold for yielding at Location 8, such that these yellow-path vehicles maintain roughly constant speeds in the three second window leading up to the red-path merge point. Due to the shorter acceleration zone at Location 5, however, these yellow-path drivers may continue to accelerate in the three seconds leading up to the red-path merge point even when congestion has slowed down the freeway, making inappropriately small gaps harder for red-path drivers to identify. If this compliance-reducing effect more than offsets the beneficial effects of congestion due to breaking drivers out of “autopilot” mode, it would explain the higher violation rates at Location 5 in the presence of congestion.

Prediction Model Results

Figure 19 compares the observed compliance rates and the predicted compliance rates for the “before” period (which is the same as that used for the larger sign analysis) and the “after” period following the reopening of the rehabilitated ramp at Location 5 in 2010. In this figure,

FIGURE 19: Observed and predicted compliance by week at Location 5
 (the “after” period refers to the weeks after ramp rehabilitation and shoulder widening).

both of our prediction models (i.e., Equations 11 and 12) are included, with error bars indicating 90% confidence intervals for the predicted values. The “before” weeks on the left side of the chart give a sense of the inherent variability of the data and the reliability of our predictions from Equations 11 and 12. We also observe that both Equations 11 and 12 generate similar predictions for the first five weeks after the rehabilitation, validating our use of Equation 12 as a substitute for Equation 11 in those weeks where Location 1 was unavailable as a control site.

The large gap in data for the “after” weeks is largely due to ramp rehabilitations at other interchanges on the corridor that either made Location 6 unavailable as a control ramp (which was needed for Equation 11 and Equation 12), or affected the performance of the experimental ramp due to diverted traffic from the westbound entrance at Del Mar Av (see Figure 3) when it was closed for rehabilitation (see Figures 4 and 5). We also ignored data from any control ramp for the first four weeks following its rehabilitation, based on the assumption that it may take about this long for conditions to return to normal at that location. As Figure 5 shows, this meant that Location 6 was unavailable as a control ramp for the four-week period between 11/14 and 12/5, despite the fact that the ramp was open during those weeks. We will revisit this assumption later in this section, when we use likelihood maximization to fit the Driver Response Curve to the observed data.

Figures 4-5 indicate the specific reasons that each missing week in the “after” period was excluded. Whenever predictions from Equation 11 were available, we used those in place of Equation 12 since the former had lower risk associated with its predictions. Although there are some weeks (i.e., 10/3 through 11/28) in the “after” period where Location 1 was available as a control ramp but Location 6 was not, other factors in those weeks had affected the behavior of drivers at Location 1 and Location 5 in ways that we could not reliably control for. Specifically, detour traffic from other adjacent ramps that were closed in those weeks (as indicated in Figure 5) was affecting driver compliance measurements at Locations 1 and 5. Consequently, there was no value in finding an alternate prediction model that did not include Location 6 but was allowed to include Location 1, since we would not have been able to apply it to any additional weeks that are not already shown in Figure 19.

Modified Driver Response Model

As Figure 19 reveals, compliance rates at the experimental ramp tended to decrease in the long term after the ramp rehabilitation was complete, rather than return to their original values once the novelty effect wore off. This was largely due to some red-path drivers learning to use the newly widened shoulder to negotiate merging with yellow-path cross traffic, rather than waiting properly at the yield point. To analyze this, we require a slightly modified version of our original Driver Response Model to properly account for this learned decrease in compliance that occurred as some drivers became aware of the wider shoulder available to them for this unintended and improper purpose.

We begin by introducing a modified version of Equation 2 that replaces the α_1 term of $R(t)$ with a new variable α'_1 , intended to capture the effect of reduced compliance over time rather than increased compliance. The fraction of drivers who are originally compliant but eventually notice the change and become noncompliant from that moment onward is given by:

$$\alpha'_1 = P(t) \cdot \beta''_a$$

Here, β''_a indicates how much the compliant group eventually shrinks after the change occurs. We make the assumption that the Awareness Curve $f_a(t_a)$ is applicable to all drivers, and not just those drivers who fall into the noncompliant category. We justify this assumption by recognizing that drivers in the compliant group are allowed to switch freely to the noncompliant group (as they have done in the case of the wider shoulder) and vice versa (as they have done in the case of the larger sign). Our new compliance rate equation becomes:

$$C(t) = P(t) - R'(t) + S'(t) \quad (13)$$

where $P(t)$ is unchanged from before, $R'(t)$ includes the new α'_1 term in place of α_1 (see Equation 3), and $S'(t)$ is a slightly simplified form of $S(t)$ (see Equation 4). This simplification occurs when we recognize that almost all drivers realize the ramp has been reopened by our first week of “after” data collection, as shown by the volumes in Figure 7. Thus, any Novelty Effect due to the superficial changes to the ramp will begin in Week 1 with no delayed Awareness effect. This is based on an assumption that the drivers will be immediately aware of the superficial changes when driving on the ramp for the first time after the month-long rehabilitation, which we expect is a sufficiently long closure to prime drivers to be prepared for new things after the ramp reopens (such as new pavement, landscaping changes, and repainted striping). When we remove the impact of delayed Awareness from the Novelty Effect term (i.e., Equation 4), we obtain the following:

$$S'(t) = \alpha_2 \int_{t_n=0}^{t_n=t} [1 - f_n(t_n)] dt_n$$

Thus, the contribution of the novelty effect in a given week (t) depends only on what percentage of Group 2 drivers has a Novelty Persistence that is at least as long as t .

The error term associated with the model predictions of Equation 13 are based on the errors associated with the uncertainty σ_{cv} in the baseline compliance prediction model, $P(t)$, and the uncertainty $\sigma_{\beta'_n}$ in the estimate for β'_n (which appears in the $S'(t)$ term). Our new formula for the approximate error in Equation 13 is:

$$\sigma_c = \sqrt{(\sigma_{c,cv})^2 + (\sigma_{c,\beta'_n})^2}$$

where

$$\sigma_{c,cv} = [1 - \beta''_a F_a(t)] \sigma_{cv}$$

$$\sigma_{c,\beta'_n} = \left[\beta'_2 \int_{t_n=0}^{t_n=t} [1 - f_n(t_n)] dt_n \right] \sigma_{\beta'_n}$$

Estimate for β'_n

Just as we did in our analysis of the larger sign at Location 8, we estimate β'_n using video footage of traffic exiting eastbound Interstate 10 to northbound New Ave during the 11 AM to 2 PM midday period, since this exit ramp is most likely to carry the same drivers as those who pass

through Location 5. Following the same procedure as that for Location 8, we estimated $\beta'_n = 0.144$ and $\sigma_{\beta'_n} = 0.0262$ from a sample of 180 drivers on this exit ramp who had the opportunity to choose between yielding properly to arterial traffic (compliant), or proceeding without waiting (noncompliant).

Preliminary Results for Driver Response to Rehabilitated Ramp

The improvement by week (per Equation 1) is shown in Figure 20, where the black dots represent the observed improvements and the error bars are 95% confidence intervals for these calculated values (the error comes from the uncertainty in $N(t)$ in Equation 1). The solid line is the overall Driver Response Curve (Equation 13) that has been fitted to the data using likelihood maximization, with the contribution from the transient Novelty Effect (i.e., Group 2 drivers) shown by a dashed line and the contribution from the Awareness Effect (i.e., drivers who became aware of the wider shoulder and gradually learned to use it to circumvent the yielding requirement) shown by a dotted line. Note that using a maximum likelihood approach to find the optimal parameter values is not equivalent to using an approach that minimizes the squared residuals, because the data points that used predictions from Equation 11 have a lower amount of uncertainty associated with them than the points that were based on Equation 12. For this reason, attempting to minimize the square residuals would be an inappropriate method of fitting the Driver Response Curve to our data set. Similar arguments can be made against a least-squares approach for the Larger Sign and Yield Line analyses as well.

As was the case in Figure 5, the curves in Figure 20 are approximate due to the fact that values for each week are dependent upon $P(t)$, which fluctuates from week to week and cannot be computed for any weeks without observed data for the covariates (e.g., weeks 6–16). In Figure 20, $P(t)$ is held at a constant value of 47.13% for all weeks after the change, which is the average $P(t)$ value over the weeks in which prediction data were available. These approximations stay within 6.6% of the true Driver Response model at all times, and were applied only for Figure 20—and not for any part of the analysis. As before, the result is a smooth Driver Response Curve that highlights improvements due to Novelty and Awareness effects, and excludes the effect of small fluctuations or gaps in the covariate data.

Preliminary Model Parameters

Because of the small data set for the “after” period (i.e., 7 weeks of data), we used our results from the larger sign analysis to make some assumptions that reduced the degrees of freedom in the Driver Response model, rather than trying to use curve-fitting techniques (i.e., likelihood maximization) to solve for all five shape parameters in addition to the forms of $f_a(t_a)$ and $f_n(t_n)$. First, we assumed that $f_a(t_a)$ and $f_n(t_n)$ had the same distributions that we found through likelihood maximization for the larger sign analysis—specifically, that $f_a(t_a)$ had a Normal distribution and $f_n(t_n)$ had a Uniform distribution. We also assumed that the proportionality constant β_3 for the Awareness Curve did not depend on the particular thing drivers were becoming aware of, such that we could apply the value of 0.83 that we identified in the larger sign analysis to our current case of the wider shoulder. However, we still expected that the shape parameters μ_a and σ_a for the Awareness curve would be different between the case of a larger sign and the case of a wider shoulder, and so we solved for optimal values of μ_a and σ_a (recall that these are connected by β_3) using data for the wider shoulder, rather than assuming the same values from the Larger Sign analysis.

FIGURE 20: Percent reduction in violation rate by week at Location 5 (ramp rehabilitation). A negative reduction for a particular week indicates that violations increased.

Based on the apparent time it took for the driver response to reach a its minimum value, we used 5 weeks as an upper limit on the mean of $f_a(t_a)$ and $f_n(t_n)$, which proved to be a non-binding constraint in both cases. The parameters that resulted in a maximum Log-Likelihood of -13.527 were:

$$\beta_a'' = 0.17$$

$$\beta_2' = 1.00$$

$$\mu_a = 0.00 \text{ weeks (accurate to } \pm 0.05)$$

$$\mu_n = 2.25 \text{ weeks (accurate to } \pm 0.05)$$

Based on these results, which are shown graphically in Figure 20, we estimate that the driver response following ramp rehabilitation at Location 5 stabilized by the fifth week, with an overall decrease in compliance of 17%. Because each ramp on this corridor was subjected to the same kind of reconstruction and rehabilitation work at about the same time (see Figures 4 and 5), we assume that these results also apply to the other locations following their rehabilitation. This gives additional justification for our decision to exclude the data from Location 6 for weeks 11/14 through 12/5, since it may take four weeks from the time this control ramp is re-opened for the driver behavior to stabilize there.

However, these results also introduce another issue: compliance at Location 6 would be expected to fall by 17% following its rehabilitation just as it did at Location 5, since the ramp shoulders were widened at both sites. The prediction model (Equations 11 and 12) uses the performance at Location 6 as it was in 2009—prior to the widening of the shoulder. Therefore, we must increase the observed compliance rates at Location 6 for the weeks following its rehabilitation (i.e., weeks 11/14 and later) before inputting those rates into Equation 11 or 12, to effectively reverse the effect of the 17% drop in compliance after 11/14 due to the wider shoulder. The formula for such an adjustment is straightforward:

$$\gamma_{orig} = \frac{\gamma_{wide}}{1 - \beta_a''} \quad (14)$$

where γ_{wide} is the observed compliance rate at Location 6 following the ramp rehabilitation, and γ_{orig} is the adjusted compliance rate to cancel out the effect of the widened shoulder.

When we make this adjustment to the compliance rates at Location 6 for the weeks after its rehabilitation, we obtain updated values for $P(t)$ according to Equation 12 (Equation 11 predictions apply to those weeks prior to the rehabilitation of Location 6). We then use likelihood maximization to solve for new Driver Response Curve parameters (β_a'' , β_2' , μ_a , and μ_n). If β_a'' has changed (or if the time to stability has changed appreciably), we apply the new value of β_a'' to Equation 14, exclude or include other weeks of data for Location 6 to ensure that we only capture those weeks where the driver response has stabilized, and solve again for the Driver Response Curve parameters that maximize likelihood. This iterative process ends when we get results for (β_a'' , β_2' , μ_a , and μ_n) that do not require any additional adjustments to γ_{orig} or any changes with respect to which weeks are considered part of the unstable period for Location 6.

Final Results for Driver Response to Rehabilitated Ramp

Figure 21 shows the final version of our Driver Response Curve (Equation 13), obtained when our iterative process converged on a value of $\beta_a'' = 0.16$. As was the case in Figure 20, note that the curves in Figure 21 are approximate but stay within 6% of the true Driver Response model at all times. As before, this approximation was applied only for Figure 21—and not for any part of the analysis—and produces a smooth Driver Response Curve that highlights improvements due to Novelty and Awareness effects, and excludes the effect of small fluctuations or gaps in the covariate data. The other curve parameters associated with this Driver Response curve are:

$$\beta_2' = 1.00$$

$$\mu_a = 0.00 \text{ weeks (accurate to } \pm 0.05)$$

$$\mu_n = 2.25 \text{ weeks (accurate to } \pm 0.05)$$

As Figure 21 illustrates, the Driver Response Curve reaches its long-term value by about week 5. The figure also shows that the right-most data point is not significantly different from 0%. However, this does not imply that the long-term value may in fact be zero, since that data point becomes significantly different from zero when we assume $\beta_a'' = 0$, as shown in Figure 20. To measure the significance of the β_a'' parameter, we apply a likelihood ratio test to a constrained version of the Driver Response Curve where β_a'' is forced to zero (such that it also becomes zero in Equation 14). From this, we obtain a p-value of 0.000, revealing that the β_a'' term makes a statistically significant contribution to the performance of the model. Note that we have one degree of freedom for this likelihood ratio test, as μ_a is already at zero in the unconstrained version of the model.

By performing similar likelihood ratio tests with other constrained versions of the Driver Response Curve, we obtained the following p-value associated with other model parameters: 0.043 for β_2' and μ_n , where the associated likelihood ratio test has two degrees of freedom. Figure 22 shows the Driver Response Curves using the parameters that maximized the likelihood function when (a) β_2' and μ_n were constrained to zero, and (b) β_a'' was constrained to zero.

Analysis of Yield Line

Because the Yield Line was installed at Location 5, which was the same location we used for our analysis of Ramp Rehabilitation and shoulder widening, we already have our prediction model (i.e., Equations 11 and 12) and our estimate for β_n' . We apply Equation 13 and the value for β_a'' (which we solved for in the Ramp Rehabilitation analysis) to the measured compliance rates for the experimental ramp and the control ramps, since all had been rehabilitated by the time the Yield Line was installed, to properly control for the effect of the wider shoulders on drivers. We also prohibit ourselves from using the control ramp data at Location 1 between 2/6 and 2/27, since these are the first four weeks after the ramp was rehabilitated and reopened, meaning that the driver behavior may be unstable during this time.

Preliminary Results

Figure 23 compares the observed and predicted compliance rates for the weeks after the Yield Line was implemented. Raw data for the experimental ramp and covariates is given in Appendix H. The “Before” period data was already plotted in Figure 19. Just as we did for the Ramp Rehabilitation analysis, the predictions in Figure 23 are based on Equations 11 and 12, with the former taking priority over the latter when both predictions were available in a given week. Gaps

FIGURE 21: Percent reduction in violation rate by week at Location 5 (ramp rehabilitation and wider shoulder), after making adjustments for the wider shoulder at Control Ramp 6 following its rehabilitation.

FIGURE 22: Re-optimized Driver Response Curves at Location 5 (ramp rehabilitation & wider shoulder) when excluding either (a) the Novelty Effect or (b) the Driver Awareness Effect.

FIGURE 23: Observed and predicted compliance by week at Location 5, for the period after the Yield Line was installed.

in the data are due to a combination of inclement weather (see Figure 5) and an expectation that bi-weekly resolution would be sufficient for the later weeks in the chart.

The improvement by week (from Equation 1) is shown in Figure 24, with the observed data represented by black dots, the Driver Response Curve predictions indicated by the lines, and the 95% confidence intervals given by the error bars. As before, the curves in Figure 24 are approximations due to the fact that the true values for each week are dependent on $P(t)$, which has a tendency to fluctuate slightly from one week to another, and is also unavailable for weeks that were not analyzed. To overcome this, we use an average value of $P(t)$ for all weeks in the “after” period for Figure 24, to smooth out the small fluctuations in $P(t)$ and to focus on the contributions of the Novelty and Awareness effects to the Driver Response Curve. The approximations have an average deviation from their true values of less than 11% at all times, and are only applied to this figure—and not to any part of the analysis.

Because we had 19 data points in our “after” period to use in solving for the parameters of the Driver Response Curve, we did not assume a particular distribution for the Novelty Persistence or Awareness effects as we had done for the Ramp Rehabilitation analysis. Instead, we compared the likelihood values achieved for each combination of candidate distributions to identify the best ones. As we did in our analysis of the Larger Sign, we tested Normal and Log-Normal distributions for the Awareness Effect, and we tested Normal, Uniform, and Geometric distributions for the Novelty Persistence Effect. We used 15 weeks as an upper limit on the mean of both distributions.

The combination that maximized the likelihood function was a Normal distribution for delayed Driver Awareness and a Normal distribution for Novelty Persistence. The optimal parameters were:

$$\beta'_a = 0.26$$

$$\beta'_2 = 1.00$$

$$\mu_a = 5.5 \text{ weeks}$$

$$\mu_n = 11.0 \text{ weeks}$$

$$\beta_3 = 0.30$$

Table 2 shows the likelihood function values across all increments of the proportionality constant β_3 , for each combination of Awareness and Novelty Persistence distributions. The optimal parameters are shown across these three variables (i.e., the value for β_3 and the distribution families for $f_a(t_a)$ and $f_n(t_n)$), as this will aid us later when we attempt to unify our separate results from the Larger Sign, Ramp Rehabilitation, and Yield Line analyses. The Log-Normal distribution tended to produce lower likelihood values than the Normal distribution for the Awareness effect (see Table 2), reinforcing the finding from the Larger Sign analysis that the Awareness effect is better modeled by a Normal distribution.

Checking our Assumption of Normally Distributed Residuals

Although the Driver Response Curve was fitted to the data using likelihood maximization, we encountered an issue regarding the optimization: the proportionality constant β_3 hit its lower bound, implying that a lower value for this parameter would have improved the model. Indeed, when we ran additional tests with lower values for β_3 , we got even higher likelihood function values with a maximum occurring around $\beta_3 = 0.10$ (see Table 2). However, as anything lower

FIGURE 24: Percent reduction in violation rate by week at Location 5 (Yield Line), using the solution from the ramp rehabilitation analysis of $\beta_a'' = 0.16$.

$f_a(t_a) : \text{Normal}$							$f_n(t_n) : \text{Normal}$						
μ_n	β'_2	μ_a	β_3	β'_a	Log Likelihood	Rank	μ_n	β'_2	μ_a	β_3	β'_a	Log Likelihood	Rank
11.5	0.95	5.5	0.05	0.26	-89.922	2	11.5	0.95	5.5	0.05	0.26	-90.123	2
11.0	1.00	6.0	0.10	0.26	-89.822	1	11.0	1.00	6.0	0.10	0.26	-89.761	1
10.5	1.00	6.0	0.15	0.26	-95.243	3	10.5	1.00	6.0	0.15	0.26	-95.067	3
10.5	1.00	6.0	0.20	0.26	-101.851	4	10.5	1.00	6.0	0.20	0.26	-101.909	4
11.0	1.00	5.5	0.25	0.26	-107.616	5	11.0	1.00	5.5	0.25	0.26	-108.881	5
11.0	1.00	5.5	0.30	0.26	-112.673	6	10.5	1.00	6.0	0.30	0.26	-114.882	6
11.0	1.00	5.5	0.35	0.26	-117.578	7	10.5	1.00	6.0	0.35	0.26	-120.122	7
11.0	1.00	5.5	0.40	0.26	-121.594	8	11.0	1.00	6.0	0.40	0.26	-125.024	8
15.0	1.00	5.0	0.45	0.22	-123.500	9	15.0	1.00	5.5	0.45	0.22	-128.876	9
15.0	1.00	5.0	0.50	0.22	-124.959	10	15.0	1.00	5.5	0.50	0.22	-131.293	10
15.0	1.00	4.5	0.55	0.22	-126.290	11	15.0	1.00	5.0	0.55	0.22	-133.374	11
15.0	1.00	4.5	0.60	0.22	-127.443	12	15.0	1.00	5.0	0.60	0.22	-135.390	12
15.0	1.00	4.5	0.65	0.22	-128.853	13	15.0	1.00	5.0	0.65	0.22	-137.400	13
15.0	1.00	4.5	0.70	0.22	-130.486	14	15.0	1.00	5.0	0.70	0.22	-139.364	14
15.0	1.00	4.0	0.75	0.22	-131.529	15	15.0	1.00	4.5	0.75	0.22	-140.994	15
15.0	1.00	4.0	0.80	0.22	-132.691	16	15.0	1.00	4.5	0.80	0.22	-142.545	16
15.0	1.00	4.0	0.85	0.22	-134.000	17	15.0	1.00	4.5	0.85	0.22	-144.017	17
15.0	1.00	4.0	0.90	0.22	-135.471	18	15.0	1.00	4.5	0.90	0.22	-145.501	18
15.0	1.00	3.5	0.95	0.22	-136.446	19	15.0	1.00	4.0	0.95	0.22	-146.862	19
11.0	1.00	5.5	0.30	0.26	-112.673	Optimal*	10.9	1.00	5.7	0.30	0.26	-114.248	Optimal*

$f_a(t_a) : \text{Normal}$							$f_n(t_n) : \text{Uniform}$						
μ_n	β'_2	μ_a	β_3	β'_a	Log Likelihood	Rank	μ_n	β'_2	μ_a	β_3	β'_a	Log Likelihood	Rank
15.0	1.00	5.5	0.05	0.22	-135.227	19	15.0	1.00	5.5	0.05	0.22	-135.442	7
15.0	1.00	6.0	0.10	0.22	-133.563	18	15.0	1.00	6.0	0.10	0.22	-133.580	4
15.0	1.00	5.5	0.15	0.22	-131.520	4	15.0	1.00	5.5	0.15	0.22	-132.479	2
15.0	1.00	5.5	0.20	0.22	-131.230	1	15.0	1.00	5.5	0.20	0.22	-132.328	1
15.0	1.00	5.5	0.25	0.22	-131.593	6	15.0	1.00	5.5	0.25	0.22	-132.941	3
15.0	1.00	5.5	0.30	0.22	-132.125	12	15.0	1.00	5.5	0.30	0.22	-133.809	5
15.0	1.00	5.5	0.35	0.22	-132.607	17	15.0	1.00	5.5	0.35	0.22	-135.248	6
15.0	1.00	5.0	0.40	0.22	-132.371	15	15.0	1.00	5.5	0.40	0.22	-136.563	8
15.0	1.00	5.0	0.45	0.22	-132.190	14	15.0	1.00	5.5	0.45	0.22	-137.724	9
15.0	1.00	5.0	0.50	0.22	-132.141	13	15.0	1.00	5.5	0.50	0.22	-138.748	10
15.0	1.00	4.5	0.55	0.22	-132.109	11	15.0	1.00	5.0	0.55	0.22	-139.491	11
15.0	1.00	4.5	0.60	0.22	-131.674	8	15.0	1.00	5.0	0.60	0.22	-139.969	12
15.0	1.00	4.5	0.65	0.22	-131.473	2	15.0	1.00	5.0	0.65	0.22	-140.440	13
15.0	1.00	4.5	0.70	0.22	-131.487	3	15.0	1.00	5.0	0.70	0.22	-140.900	14
15.0	1.00	4.5	0.75	0.22	-131.703	9	15.0	1.00	5.0	0.75	0.22	-141.348	15
15.0	1.00	4.0	0.80	0.22	-131.565	5	15.0	1.00	5.0	0.80	0.22	-141.782	16
15.0	1.00	4.0	0.85	0.22	-131.617	7	15.0	1.00	5.0	0.85	0.22	-142.201	17
15.0	1.00	4.0	0.90	0.22	-131.937	10	15.0	1.00	5.0	0.90	0.22	-142.606	18
15.0	1.00	4.0	0.95	0.22	-132.398	16	15.0	1.00	4.5	0.95	0.22	-142.977	19
15.5	1.00	4.4	0.70	0.23	-130.599	Optimal*	15.5	1.00	5.5	0.30	0.22	-133.220	Optimal*

*Note that the optimal value of $B3$ is taken from the range of 0.30–1.00 only, since values outside this range were deemed unrealistic.

TABLE 2 (Part 1 of 2): Optimal parameters and associated Log-Likelihoods for each combination of β_3 , $f_a(t_a)$, and $f_n(t_n)$ at Location 5 (for the driver response curves for the Yield Line). Higher-resolution overall optimal values are given in the last row of each table.

$f_a(t_a) : \text{Normal}$				$f_n(t_n) : \text{Geometric}$			$f_a(t_a) : \text{Log-Normal}$				$f_n(t_n) : \text{Geometric}$		
μ_n	β'_2	μ_a	β_3	β'_a	Log Likelihood	Rank	μ_n	β'_2	μ_a	β_3	β'_a	Log Likelihood	Rank
0.07	1.00	5.5	0.05	0.22	-147.859	17	0.07	1.00	6.0	0.05	0.22	-147.953	6
0.07	1.00	6.0	0.10	0.22	-144.872	4	0.07	1.00	6.0	0.10	0.22	-144.894	1
0.07	1.00	5.5	0.15	0.22	-144.506	2	0.07	1.00	6.0	0.15	0.22	-144.917	2
0.07	1.00	5.5	0.20	0.22	-144.339	1	0.07	1.00	5.5	0.20	0.22	-145.531	3
0.07	1.00	5.5	0.25	0.26	-144.631	3	0.07	1.00	5.5	0.25	0.26	-146.311	4
0.07	1.00	5.5	0.30	0.26	-145.005	5	0.07	1.00	5.5	0.30	0.26	-147.327	5
0.07	1.00	5.5	0.35	0.26	-145.618	6	0.07	1.00	6.0	0.35	0.26	-148.828	7
0.07	1.00	5.5	0.40	0.26	-146.267	9	0.07	1.00	6.0	0.40	0.26	-150.183	8
0.07	1.00	5.0	0.45	0.26	-146.355	10	0.07	1.00	5.5	0.45	0.26	-151.665	9
0.07	1.00	5.0	0.50	0.26	-146.156	8	0.07	1.00	5.5	0.50	0.26	-152.969	10
0.07	1.00	5.0	0.55	0.26	-146.151	7	0.07	1.00	5.5	0.55	0.26	-154.186	11
0.07	1.00	5.0	0.60	0.26	-146.361	11	0.07	1.00	5.5	0.60	0.26	-155.299	12
0.07	1.00	4.5	0.65	0.26	-146.495	13	0.07	1.00	5.5	0.65	0.26	-156.314	13
0.07	1.00	4.5	0.70	0.26	-146.401	12	0.07	1.00	5.5	0.70	0.26	-157.240	14
0.07	1.00	4.5	0.75	0.26	-146.558	14	0.07	1.00	5.5	0.75	0.26	-158.090	15
0.07	1.00	4.5	0.80	0.26	-146.962	15	0.07	0.95	7.0	0.80	0.30	-178.535	19
0.07	1.00	4.5	0.85	0.26	-147.639	16	0.07	0.95	6.5	0.85	0.30	-175.364	18
0.07	1.00	4.0	0.90	0.26	-148.034	18	0.07	0.95	5.5	0.90	0.30	-173.742	17
0.07	1.00	4.0	0.95	0.26	-148.480	19	0.07	1.00	6.5	0.95	0.30	-172.389	16
0.10	1.00	5.4	0.30	0.24	-143.437	Optimal*	0.10	1.00	5.6	0.30	0.24	-145.395	Optimal*

*Note that the optimal value of $B3$ is taken from the range of 0.30–1.00 only, since values outside this range were deemed unrealistic.

TABLE 2 (Part 2 of 2): Optimal parameters and associated Log-Likelihoods for each combination of β_3 , $f_a(t_a)$, and $f_n(t_n)$ at Location 5 (for the driver response curves for the Yield Line). Higher-resolution overall optimal values are given in the last row of each table.

than 0.3 for β_3 makes the Driver Awareness distribution unreasonably narrow, we were reluctant to relax this constraint despite the apparent improvement we would achieve in our model.

Instead, we revisited one of our assumptions of the likelihood maximizing approach to solving for the curve parameters (β'_a , β'_2 , μ_a , μ_n , and β_3). Originally, we had assumed that the standard scores (sometimes called “Z-scores”) for each residual value followed a Standard Normal distribution. However, for a sample of only 19 points, some of our residuals had unusually high standard scores; for example, the three most extreme residuals had standard scores of -6.28, -5.44, and 5.24). We can use a Q-Q plot to evaluate how well a Standard Normal Distribution fits our data (i.e., the standard scores of the residuals from our optimized Driver Response Curve). For such a Q-Q plot, the observed quantiles for our set of data points are plotted against the theoretical quantiles for a particular distribution that we want to compare to (in this case, a Normal distribution). If the resultant plot tracks closely with the $y = x$ line, it suggests that the observed data follows the distribution of comparison.

Figure 25b shows the Q-Q plot comparing the quantiles of a Standard Normal distribution to the quantiles of the standard scores of the residuals from our best-fit Driver Response Curve. As an additional check, we also generate a similar Q-Q plot (shown in Figure 25a) using the residuals from our best-fit Driver Response Curve for the Larger Sign analysis. For comparative purposes, Figure 25c also includes a Q-Q plot comparing the quantiles of a Standard Normal distribution to the quantiles of a randomly-generated set of 19 data points that were drawn from a Standard Normal distribution, to give a sense of how closely the data would typically fit the $y = x$ line if the two distributions being compared were a match.

Figure 25 reveals that although the Standard Normal distribution assumption for the residuals of our Driver Response curve held up in our analysis of the Larger Sign, it appears to be an inappropriate assumption for our current analysis of the Yield Line. We observe, however, that the plotted points for the standard scores of the Yield Line residuals fall along a line that passes through the origin, implying a linear relationship between a Standard Normal distribution and the underlying distribution of our residual standard scores. When we reduce all of our residual standard scores by a factor of 3, we find a strong fit between the residual data and the theoretical distribution, as shown in Figure 25d. This suggests that our original estimates for the uncertainty associated with each point were underestimated by a factor of 3, given the assumption that they are appropriately modeled by a Driver Response Curve in the form of Equation 2 (just as the Larger Sign data was).

One possible source of such an underestimate could be inconsistencies in the performance of a research assistant in evaluating the compliance rates at a particular location from week to week, since our method of calibrating the assistants’ results to ours was based on scale factors that assumed constant performance throughout all the weeks of data that a particular assistant did. Another possible source would be the speed data for the freeway, which had to be estimated some weeks by watching the video footage, but was obtained in other weeks using loop detectors in the freeway. The majority of the speed data from the “before” period, which was used to estimate the coefficients of Equations 11 and 12 and to evaluate the prediction risk of those models, came from loop detectors, whereas the majority of the speed data for the “after” period had to be obtained from sampling vehicles on the freeway using the video data. A similar situation arose for the on-ramp volume data, since it had to be manually counted in those weeks when the on-ramp detector was unavailable (which largely occurred in the “after” period only).

(a) Q-Q Plot for Larger Sign (Location 8).

(b) Q-Q Plot for Yield Line (Location 5).

(c) Q-Q Plot for 19 points randomly drawn from a Normal Distribution.

(d) Q-Q Plot for Yield Line (Location 5) with Z-scores for observed data reduced by a factor of 3.

FIGURE 25: Q-Q Plots comparing the quantiles of the standard scores of the residuals from our Driver Response Curves to the theoretical quantiles for a Standard Normal distribution. The diagonal line across each plot is the $y = x$ line.

Final Results

In an attempt to fit a more accurate Driver Response Curve to our observed data for the Yield Line, we now repeat the curve fitting procedure with the assumption that the errors associated with each data point have been underestimated by a factor of approximately 3. In doing so, we effectively reduce the standard scores associated with each residual by a factor of 3. We then repeat our likelihood maximization procedure to find the shape parameters that produce the best fit between the Driver Response Curve and the observed data. Our results are shown in Table 3.

The results are very similar to those obtained in the first round of likelihood maximization shown in Table 2. This is not unexpected, but has the undesirable consequence that our optimal β_3 value is still outside the expected range of 0.3 to 1.0. We therefore enforce a lower bound on β_3 to keep our results reasonable. Our original range of β_3 values (i.e., from 0.3 to 1.0) was based on literature that suggested an awareness distribution proportionality constant of approximately 0.5. We now seek to establish a more precise lower limit based on a closer examination of the available literature.

In one frequently cited study of driver response times to an amber signal at an intersection, the author estimated a mean of 1.3 to 1.4 seconds and a standard deviation of 0.55 to 0.75, suggesting a lower limit of $\beta_3 = 0.392$ (11). In a separate meta-analysis of driver reaction times, another researcher found that driver response time to an unexpected, but common, event (e.g., slow traffic ahead) had a mean of between 1.2 and 1.35 seconds with an estimated standard deviation of 0.6, suggesting a lower limit of $\beta_3 = 0.44$ (10). Finally, we obtain additional insight from a study that measured perception-reaction times based on the release of the accelerator rather than depressing the brake pedal, to better capture the perception portion of the distribution. In this study, the authors reported a mean of 0.4859 seconds and a standard deviation of 0.1538 seconds for drivers' responses to an amber signal, and a mean of 0.70 and a standard deviation of 0.316 for responses to an unexpected warning (either infrastructure-based or in-vehicle), suggesting a lower limit $\beta_3 = 0.313$ (12).

As discussed earlier in the section on "Modeling the Driver Response over Time," these studies of driver reaction are not perfect analogs to our current analysis of driver awareness to changes in the environment (such as a recently-added yield line), but it can still offer insights into the reasonable distribution for how long each of these things may take. Given this range of potential values for β_3 , we take 0.3 as a reasonable lower bound and arrive at the following maximum likelihood estimates for the Driver Response Curve parameters:

$$\beta'_a = 0.26$$

$$\beta'_2 = 1.00$$

$$\mu_a = 5.5 \text{ weeks}$$

$$\mu_n = 10.9 \text{ weeks}$$

$$\beta_3 = 0.30$$

Our value for β'_2 hit its upper bound during likelihood maximization just as it did for the Larger Sign analysis. However, this active constraint on β'_2 should not have a significant effect on our evaluation of the long-term effectiveness of the Yield Line, since our MOE is independent of β'_2 . The optimal values for β'_2 for the Yield Line analysis and Larger Sign analysis suggest that the proportion of drivers that are deliberately noncompliant at these entrance ramps is higher than

$f_a(t_a) : \text{Normal}$							$f_n(t_n) : \text{Normal}$						
μ_n	β'_2	μ_a	β_3	β'_a	Log Likelihood	Rank	μ_n	β'_2	μ_a	β_3	β'_a	Log Likelihood	Rank
11.5	0.95	5.5	0.05	0.26	-17.086	2	11.5	0.95	5.5	0.05	0.26	-17.119	2
10.5	1.00	6.0	0.10	0.26	-16.872	1	10.5	1.00	6.0	0.10	0.26	-16.848	1
10.5	1.00	6.0	0.15	0.26	-17.687	3	10.5	1.00	6.0	0.15	0.26	-17.593	3
10.5	1.00	6.0	0.20	0.26	-18.621	4	10.5	1.00	6.0	0.20	0.26	-18.651	4
11.0	1.00	5.5	0.25	0.26	-19.366	5	11.0	1.00	5.5	0.25	0.26	-19.505	5
11.0	1.00	5.5	0.30	0.26	-20.084	6	11.0	1.00	5.5	0.30	0.26	-20.478	6
11.0	1.00	5.5	0.35	0.26	-20.931	7	11.0	1.00	5.5	0.35	0.26	-21.430	7
11.0	1.00	5.5	0.40	0.26	-21.598	8	10.5	1.00	6.0	0.40	0.26	-22.139	8
15.0	1.00	5.0	0.45	0.22	-21.969	9	15.0	1.00	5.5	0.45	0.22	-22.836	9
15.0	1.00	4.5	0.50	0.22	-22.121	10	15.0	1.00	5.5	0.50	0.22	-23.144	10
15.0	1.00	4.5	0.55	0.22	-22.203	11	15.0	1.00	5.0	0.55	0.22	-23.393	11
15.0	1.00	4.5	0.60	0.22	-22.308	12	15.0	1.00	5.0	0.60	0.22	-23.601	12
15.0	1.00	4.5	0.65	0.22	-22.432	13	15.0	1.00	4.5	0.65	0.22	-23.788	13
15.0	1.00	4.0	0.70	0.22	-22.496	14	15.0	1.00	4.5	0.70	0.22	-23.922	14
15.0	1.00	4.0	0.75	0.22	-22.568	15	15.0	1.00	4.5	0.75	0.22	-24.049	15
15.0	1.00	4.0	0.80	0.22	-22.657	16	15.0	1.00	4.0	0.80	0.22	-24.136	16
15.0	1.00	4.0	0.85	0.22	-22.760	17	15.0	1.00	4.0	0.85	0.22	-24.214	17
15.0	1.00	3.5	0.90	0.22	-22.878	18	15.0	1.00	4.0	0.90	0.22	-24.284	18
15.0	1.00	3.5	0.95	0.22	-22.967	19	15.0	1.00	4.0	0.95	0.22	-24.400	19
10.9	1.00	5.5	0.30	0.26	-20.078	Optimal*	10.9	1.00	5.6	0.30	0.26	-20.394	Optimal*

$f_a(t_a) : \text{Normal}$							$f_n(t_n) : \text{Uniform}$						
μ_n	β'_2	μ_a	β_3	β'_a	Log Likelihood	Rank	μ_n	β'_2	μ_a	β_3	β'_a	Log Likelihood	Rank
15.0	1.00	5.5	0.05	0.22	-23.272	19	15.0	1.00	5.5	0.05	0.22	-23.312	6
15.0	1.00	6.0	0.10	0.22	-22.989	18	15.0	1.00	6.0	0.10	0.22	-22.990	4
15.0	1.00	5.5	0.15	0.22	-22.559	3	15.0	1.00	5.5	0.15	0.22	-22.751	1
15.0	1.00	5.5	0.20	0.22	-22.586	5	15.0	1.00	5.5	0.20	0.22	-22.796	2
15.0	1.00	5.5	0.25	0.22	-22.700	9	15.0	1.00	5.5	0.25	0.22	-22.922	3
15.0	1.00	5.5	0.30	0.22	-22.823	12	15.0	1.00	5.5	0.30	0.22	-23.057	5
15.0	1.00	5.5	0.35	0.22	-22.916	14	15.0	1.00	5.5	0.35	0.22	-23.360	7
15.0	1.00	5.0	0.40	0.22	-22.941	15	15.0	1.00	5.5	0.40	0.22	-23.614	8
15.0	1.00	5.0	0.45	0.22	-22.947	17	15.0	1.00	5.5	0.45	0.22	-23.813	9
15.0	1.00	5.0	0.50	0.22	-22.946	16	15.0	1.00	5.5	0.50	0.22	-23.968	10
15.0	1.00	4.5	0.55	0.22	-22.869	13	15.0	1.00	5.5	0.55	0.22	-24.085	11
15.0	1.00	4.5	0.60	0.22	-22.791	11	15.0	1.00	5.0	0.60	0.22	-24.129	15
15.0	1.00	4.5	0.65	0.22	-22.734	10	15.0	1.00	5.0	0.65	0.22	-24.157	18
15.0	1.00	4.0	0.70	0.22	-22.695	8	15.0	1.00	4.5	0.70	0.22	-24.174	19
15.0	1.00	4.0	0.75	0.22	-22.607	6	15.0	1.00	4.5	0.75	0.22	-24.156	17
15.0	1.00	4.0	0.80	0.22	-22.543	2	15.0	1.00	4.5	0.80	0.22	-24.141	16
15.0	1.00	4.0	0.85	0.22	-22.533	1	15.0	1.00	4.5	0.85	0.22	-24.126	14
15.0	1.00	4.0	0.90	0.22	-22.575	4	15.0	1.00	4.5	0.90	0.22	-24.112	13
15.0	1.00	4.0	0.95	0.22	-22.629	7	15.0	1.00	4.5	0.95	0.22	-24.098	12
15.5	1.00	4.1	0.82	0.23	-22.346	Optimal*	15.5	1.00	5.5	0.30	0.22	-22.930	Optimal*

*Note that the optimal value of $B3$ is taken from the range of 0.30–1.00 only, since values outside this range were deemed unrealistic.

TABLE 3 (Part 1 of 2): Optimal parameters and associated Log-Likelihoods for each combination of β_3 , $f_a(t_a)$, and $f_n(t_n)$ at Location 5 (for the driver response curves for the Yield Line), after adjusting the error terms by a factor of 3.
Higher-resolution overall optimal values are given in the last row of each table.

$f_a(t_a) : \text{Normal}$				$f_n(t_n) : \text{Geometric}$			$f_a(t_a) : \text{Log-Normal}$				$f_n(t_n) : \text{Geometric}$		
μ_n	β'_2	μ_a	β_3	β'_a	Log Likelihood	Rank	μ_n	β'_2	μ_a	β_3	β'_a	Log Likelihood	Rank
0.07	1.00	5.5	0.05	0.22	-24.587	14	0.07	1.00	5.5	0.05	0.22	-24.630	6
0.07	1.00	6.0	0.10	0.22	-24.244	4	0.07	1.00	6.0	0.10	0.22	-24.245	3
0.07	1.00	5.5	0.15	0.22	-23.974	1	0.07	1.00	5.5	0.15	0.22	-24.171	1
0.07	1.00	5.5	0.20	0.22	-24.021	2	0.07	1.00	5.5	0.20	0.22	-24.243	2
0.07	1.00	5.5	0.25	0.22	-24.168	3	0.07	1.00	5.5	0.25	0.22	-24.408	4
0.07	1.00	5.5	0.30	0.22	-24.337	5	0.07	1.00	5.5	0.30	0.22	-24.587	5
0.07	1.00	5.5	0.35	0.22	-24.481	6	0.07	1.00	5.5	0.35	0.22	-24.906	7
0.07	1.00	5.0	0.40	0.22	-24.583	13	0.07	1.00	5.5	0.40	0.22	-25.208	8
0.07	1.00	5.0	0.45	0.22	-24.629	16	0.07	1.00	5.5	0.45	0.22	-25.461	9
0.07	1.00	5.0	0.50	0.22	-24.668	18	0.07	1.00	5.5	0.50	0.22	-25.753	10
0.07	1.00	4.5	0.55	0.22	-24.685	19	0.07	1.00	5.5	0.55	0.26	-25.937	11
0.07	1.00	4.5	0.60	0.22	-24.639	17	0.07	1.00	5.5	0.60	0.26	-26.059	12
0.07	1.00	4.5	0.65	0.26	-24.572	12	0.07	1.00	5.5	0.65	0.26	-26.154	13
0.07	1.00	4.5	0.70	0.26	-24.527	10	0.07	1.00	5.0	0.70	0.26	-26.205	14
0.07	1.00	4.5	0.75	0.26	-24.504	8	0.07	1.00	5.0	0.75	0.26	-26.238	15
0.07	1.00	4.0	0.80	0.26	-24.492	7	0.07	1.00	5.0	0.80	0.26	-26.262	16
0.07	1.00	4.0	0.85	0.26	-24.514	9	0.07	1.00	5.0	0.85	0.26	-26.281	17
0.07	1.00	4.0	0.90	0.26	-24.552	11	0.07	1.00	5.0	0.90	0.26	-26.294	18
0.07	1.00	4.0	0.95	0.26	-24.607	15	0.07	1.00	5.0	0.95	0.26	-26.310	19
0.10	1.00	5.3	0.30	0.23	-24.013	Optimal*	0.10	1.00	5.5	0.30	0.23	-24.381	Optimal*

*Note that the optimal value of $B3$ is taken from the range of 0.30–1.00 only, since values outside this range were deemed unrealistic.

TABLE 3 (Part 2 of 2): Optimal parameters and associated Log-Likelihoods for each combination of β_3 , $f_a(t_a)$, and $f_n(t_n)$ at Location 5 (for the driver response curves for the Yield Line), after adjusting the error terms by a factor of 3.
Higher-resolution overall optimal values are given in the last row of each table.

that for the stop signs based on the way we have defined it in either case, despite the similarities between the two situations as discussed in the section on “Estimating β'_n .”

A look at the Log-Likelihood values in Table 3 reveals that the Log-Normal distribution for the Awareness effect typically did worse than the Normal distribution at characterizing the trends in the compliance data across all three candidate distributions for Novelty Persistence. This finding is consistent with our results from the Larger Sign analysis. One difference, however, is that a Normal distribution for $f_n(t_n)$ provides a slightly better fit to the data than a Uniform distribution does, whereas the reverse was true for the Larger Sign data. In the “Unified Results” section later, we will attempt to combine the likelihoods of each combination of $f_n(t_n)$, $f_a(t_a)$, and β_3 from Tables 1 and 3 to arrive at a unified, consistent set of results for both compliance strategies.

Our estimated long-term effectiveness of the Yield Line was 26% according to Equation 6, with the Driver Response Curve taking approximately 25 weeks to stabilize as shown in Figure 26. The Novelty Effect alone was found to have a mean persistence of 10 weeks, which is twice as long as the estimated mean for a Larger Sign. This is reasonable given that the Yield Lines are a relatively new addition to the MUTCD and are less prevalent than large Yield signs, making it likely that an encounter with a Yield Line would be a more novel occurrence for drivers than coming upon a large Yield sign. This also supports our finding from the Larger Sign analysis that it takes longer for the driver response to stabilize than the 4-8 weeks often assumed in literature. In the “Sensitivity Analysis” section later, we will use likelihood ratios to develop a confidence interval for this MOE estimate as well as our MOE for the larger Yield sign.

We estimate that it took drivers an average of 5 weeks to notice, interpret, and respond to the Yield Line, which is comparable to the mean duration of 3.7 weeks for the Larger Sign. Although we anticipated a greater degree of visibility for the Yield Line due to its placement on the pavement—where it has fewer distractions to compete with than a sign at eye level—the slightly longer response time for the Yield Line compared to the larger sign may be due to a more extended “learning” period associated with it. Being uncommon and less familiar to drivers, it is reasonable to expect that it would take longer for drivers to properly interpret the meaning of a Yield Line.

The improvement data for the Yield Line exhibits a trend of initial growth in the weeks following its implementation, which supports the results we obtained in our Larger Sign analysis. This runs contrary to the existing literature, which commonly assumes that the improvement in compliance starts off at some peak value immediately after the implementation, and then gradually tapers off to a lower steady-state value over time. This revealed trend of increasing compliance in the first few weeks has important consequences for reliably estimating a strategy’s effectiveness, as we explore later in the section titled, “Comparison to Traditional Before-and-After Approach.”

Significance of Parameters

Just as we did for the Larger Sign analysis, we now examine the significance of various model parameters by constraining them to zero and performing Likelihood Ratio Tests. From this procedure, we obtain a p-value of 0.0000 for μ_a , providing compelling evidence that the Driver Awareness effect exists. Note that unlike the case of the Larger Sign, the Chi-squared statistic used for this test involves only one degree of freedom, since β_3 is still relevant in Equation 2 even when μ_a is constrained to zero (because $f_n(t_n)$ follows a Normal distribution in this case, rather than a Uniform distribution as it did in our Larger Sign analysis).

FIGURE 26: Percent reduction in violation rate by week at Location 5 (Yield Line), using rescaled error terms for the data points when performing the likelihood maximization.

A similar test on the significance of μ_n and β'_2 returns a p-value of 0.0002, which is strong evidence that the Novelty Effect is also a significant contributor to the driver response following the implementation of a compliance strategy. Although the strength of the Novelty Effect was not found to be statistically significant on a 10% level in the case of the Larger Sign, our data for the Yield Line shows significance even at a 1% level. For this test, we have two degrees of freedom for the Chi-squared statistic, since μ_n cannot exist in the Driver Response Curve without β'_2 , and vice versa (see Equation 2). Figure 27 shows the Driver Response Curves using the parameters that maximized the likelihood function when either μ_a was constrained to zero or when β'_2 and μ_n were constrained to zero, to illustrate the significance of their effects on the Driver Response Curve.

6. UNIFIED RESULTS, SENSITIVITY ANALYSIS, AND ALTERNATE APPROACHES

Now that we have estimates for the parameters of the Driver Response Curve for both a Yield Line and a larger sign, we attempt to reconcile some of the differences between our two models by arriving at a single estimate for β_3 , as well as a single distribution family for $f_n(t_n)$. To do this, we combine the likelihood values for each combination of $f_n(t_n)$ and β_3 , and search for the set that maximizes the overall likelihood. Table 4 shows the composite likelihood values associated with each combination, which are calculated by taking the product of the component likelihood results from the Larger Sign and Yield Line analyses (see Tables 1 and 3).

In searching for the optimal combination, we began by combining the likelihood values obtained from the first pass of the likelihood maximization procedure for each compliance strategy (i.e., the likelihood values that arose from the brute force optimizations using the coarse increments described in the “Driver Response Curve” section). With these results, we then identified three candidate optima:

- a) $\beta_3 = 0.30$ with a Normal distribution for $f_n(t_n)$
- b) $\beta_3 = 0.80$ with a Normal distribution for $f_n(t_n)$
- c) $\beta_3 = 0.85$ with a Uniform distribution for $f_n(t_n)$

Next, we conducted a second pass using higher-resolution increments for all of the shape parameters (i.e., β'_a , β'_2 , μ_a , μ_n , and β_3) for the Larger Sign and the Yield Line, and combined the associated likelihood values for each compliance strategy. After this second pass in our search for optimal $f_n(t_n)$ and β_3 , we obtained more precise estimates for the optima listed above, as well as their associated Log-likelihood values:

- a) $\beta_3 = 0.30$ with a Normal distribution for $f_n(t_n)$ and a Log-Likelihood of -37.654
- b) $\beta_3 = 0.75$ with a Normal distribution for $f_n(t_n)$ and a Log-Likelihood of -37.828
- c) $\beta_3 = 0.82$ with a Uniform distribution for $f_n(t_n)$ and a Log-Likelihood of -37.441

From the Log-likelihood values, we conclude that the optimal combination is given by (c). However, in the case of the Yield Line, the parameters associated with this unified optimal solution include a Novelty Persistence mean that hits its upper bound. If left unconstrained, the likelihood maximization would continue to push the Novelty mean higher and the β'_a term lower. In fact, the unconstrained optimum has a value for μ_n well above 100 weeks, which violates our treatment of the Novelty effect (and the characterization of the Novelty effect in literature) as only a temporary effect. This makes solution (c) a bit unrealistic with respect to the Driver Response Curve for the Yield Line, and motivates us to consider the next-ranked option instead: solution (a). Because solution (a) specifies the distribution family for $f_n(t_n)$ and the value for β_3 that we were already using for the optimal Driver Response Curve for the Yield Line data, the

FIGURE 27: Re-optimized Driver Response Curves at Location 5 (Yield Line) when excluding either (a) the Driver Awareness Effect or (b) the Novelty Effect.

$f_a(t_a) : \text{Normal}$					$f_n(t_n) : \text{Normal}$				
β_3	Log Likelihood (Larger Sign)	Log Likelihood (Yield Line)	Combined Log Likelihood	Rank	β_3	Log Likelihood (Larger Sign)	Log Likelihood (Yield Line)	Combined Log Likelihood	Rank
0.30	-17.875	-20.084	-37.959	1	0.30	-17.875	-22.823	-40.698	14
0.35	-17.455	-20.931	-38.386	7	0.35	-17.455	-22.916	-40.371	13
0.40	-17.098	-21.598	-38.697	10	0.40	-17.098	-22.941	-40.039	12
0.45	-16.829	-21.969	-38.798	13	0.45	-16.829	-22.947	-39.776	11
0.50	-16.635	-22.121	-38.757	11	0.50	-16.635	-22.946	-39.581	10
0.55	-16.574	-22.203	-38.777	12	0.55	-16.677	-22.869	-39.546	9
0.60	-16.317	-22.308	-38.625	8	0.60	-16.332	-22.791	-39.123	8
0.65	-15.929	-22.432	-38.361	6	0.65	-15.917	-22.734	-38.651	7
0.70	-15.782	-22.496	-38.277	3	0.70	-15.719	-22.695	-38.415	6
0.75	-15.745	-22.568	-38.313	4	0.75	-15.608	-22.607	-38.216	4
0.80	-15.570	-22.657	-38.228	2	0.80	-15.425	-22.543	-37.968	2
0.85	-15.555	-22.760	-38.315	5	0.85	-15.388	-22.533	-37.921	1
0.90	-15.773	-22.878	-38.651	9	0.90	-15.465	-22.575	-38.040	3
0.95	-15.997	-22.967	-38.965	14	0.95	-15.608	-22.629	-38.237	5
<i>High Resolution Analysis Around $\beta_3 = 0.30$</i>					<i>High Resolution Analysis Around $\beta_3 = 0.85$</i>				
0.25	-17.799	-19.320	-37.118	1	0.80	-15.110	-22.364	-37.474	4
0.26	-17.771	-19.470	-37.241	2	0.81	-15.101	-22.354	-37.455	3
0.27	-17.725	-19.584	-37.309	3	0.82	-15.095	-22.346	-37.441	1
0.28	-17.677	-19.745	-37.422	4	0.83	-15.093	-22.350	-37.443	2
0.29	-17.628	-19.905	-37.533	5	0.84	-15.125	-22.357	-37.481	5
0.30	-17.577	-20.078	-37.654	6	0.85	-15.162	-22.365	-37.527	6
0.31	-17.524	-20.258	-37.782	7	0.86	-15.202	-22.373	-37.575	7
0.32	-17.471	-20.436	-37.907	8	0.87	-15.244	-22.382	-37.625	8
0.33	-17.417	-20.607	-38.024	9	0.88	-15.288	-22.390	-37.678	9
0.34	-17.362	-20.764	-38.127	10	0.89	-15.330	-22.397	-37.727	10
0.35	-17.308	-20.913	-38.221	11	0.90	-15.372	-22.405	-37.777	11
<i>High Resolution Analysis Around $\beta_3 = 0.80$</i>									
0.75	-15.408	-22.420	-37.828	1					
0.76	-15.393	-22.443	-37.837	2					
0.77	-15.383	-22.465	-37.848	3					
0.78	-15.393	-22.478	-37.871	4					
0.79	-15.383	-22.492	-37.875	5					
0.80	-15.390	-22.504	-37.894	6					
0.81	-15.402	-22.515	-37.917	7					
0.82	-15.415	-22.528	-37.943	8					
0.83	-15.423	-22.540	-37.963	9					
0.84	-15.435	-22.554	-37.988	10					
0.85	-15.474	-22.567	-38.041	11					

TABLE 4: Combined Log-Likelihoods from the Driver Response Curves of the Larger Sign and Yield Line, for each value of β_3 and different forms of $f_n(t_n)$. The top section of each table is used to identify candidate optimal values of β_3 on a coarse level for each $f_n(t_n)$, with higher-resolution optimal values identified in the lower sections.

parameters for it are the same as those reported in the “Final Results” section of the Yield Line analysis earlier. For solution (a), the remaining parameters of the Driver Response Curve for the Larger Sign are:

$$\beta'_a = 0.22$$

$$\beta'_2 = 0.00$$

$$\mu_a = 1.6 \text{ weeks}$$

The only active constraints in this solution are for β'_2 . In the case of the Larger Sign, β'_2 hits its lower bound, implying that when $\beta_3 = 0.30$, the best fit Driver Response Curve is obtained when no Novelty Effect is included. In the case of the Yield Line, β'_2 hits its upper bound, implying that we have underestimated the value of β'_n at this ramp (as we discussed briefly in the earlier section, “Yield Line Results”). Note that in the case of the larger sign, there is no μ_n parameter specified. This is due to the value of zero for β'_2 , which makes the μ_n term irrelevant. Since both of these active constraints are not cause for concern with respect to our optimization (i.e., they can be interpreted in a reasonable way—specifically, they suggest that no Novelty Effect is present), we conclude that this is the best overall combination for β_3 and $f_n(t_n)$ based on our data for the Yield Line and the larger Yield sign. Figure 28 shows the new Driver Response Curve for the Larger Sign data based on a value of $\beta_3 = 0.30$ and Normal distribution for $f_n(t_n)$. The Driver Response Curve for the Yield Line data is unchanged from Figure 26.

In Table 4, we have only tested a Uniform distribution and Normal distribution for $f_n(t_n)$, since a Geometric distribution was previously found to be the worst performer of the three in each compliance strategy analysis. Furthermore, we do not test different distributions for $f_a(t_a)$, since our results from the Larger Sign and the Yield Line analyses both suggest that it follows a Normal distribution (and not a Log-Normal distribution).

Revisiting the Ramp Rehabilitation and Yield Line Analyses

Now that we have identified an optimal value for β_3 and an optimal distribution for $f_n(t_n)$ based on the data for both the Yield Line and the larger Yield sign, we return to our original analysis of the Ramp Rehabilitation at Location 5 with these new findings, which we will refer to as our “initial unified results.” Recall that the original analysis of the Ramp Rehabilitation data used a β_3 of 0.83 and a Uniform distribution for $f_n(t_n)$, which was based on the results for the Larger Sign data alone. With our unified results, we now repeat our analysis of the data for the Ramp Rehabilitation using our new value for β_3 and our new distribution for $f_n(t_n)$, following the procedure outlined in the “Analysis of Ramp Rehabilitation” section earlier. From this, we obtain a revised estimate for β''_a that is based on both the Larger Sign and Yield Line data sets, rather than just the former.

With this updated value of β''_a , we then repeat our analysis of the Yield Line data to obtain a revised estimate of its Driver Response Curve parameters. We let β_3 remain unconstrained during the likelihood maximization, but still enforce a lower bound of 0.30 as justified earlier in the “Final Results” section of the Yield Line analysis. This differs from our approach for the Ramp Rehabilitation analysis, where we forced β_3 to 0.30 because the data set was inappropriately small to solve for five separate shape parameters simultaneously.

From these new Yield Line results, we create an updated set of composite likelihood values (analogous to those shown in Table 4) that draw from the data for both compliance

FIGURE 28: Percent reduction in violation rate by week at Location 8 (larger sign) using the unified solution specifications of 0.30 for β_3 and a Normal distribution for $f_n(t_n)$.

strategies. With these combined likelihoods, we identify a new unified optimal value for β_3 and distribution family for $f_n(t_n)$, which we will call our “final” results for the iteration. If these results for β_3 and $f_n(t_n)$ agree with the initial unified results mentioned earlier, we are finished. Otherwise, we repeat the process described in this section with the final β_3 and $f_n(t_n)$ results from the previous iteration becoming the new initial unified results for the next iteration, and stop when the initial and final results for β_3 and $f_n(t_n)$ agree in a particular iteration.

Updated Results for the Ramp Rehabilitation and Yield Line

Our updated Driver Response Curve for the Ramp Rehabilitation analysis using a Normal distribution for $f_n(t_n)$ and a value of 0.30 for β_3 is given in Figure 29. The shape parameters that maximized the likelihood function were:

$$\beta_a'' = 0.17$$

$$\beta_2' = 0.85$$

$$\mu_a = 0.00 \text{ weeks (accurate to } \pm 0.05)$$

$$\mu_n = 2.80 \text{ weeks (accurate to } \pm 0.05)$$

The associated Log-Likelihood value for this combination of parameters was -14.457. The parameter values are similar to those of the original results, with stability reached by week 5 as before. We also obtain a p-value associated with constraining β_a'' to zero of 0.000, revealing that the β_a'' term still makes a significant contribution to the model. Therefore, we will use it to make appropriate adjustments to the rehabilitated ramp data used in the Yield Line analysis according to Equation 14—just as we did before.

For our updated Yield Line analysis, we again restrict ourselves from using the control ramp data at Location 1 between 2/6 and 2/27 due to the expected four-week period of instability following the rehabilitation of that ramp. This is based on the duration of the unstable period revealed in Figure 29. Our updated Driver Response Curve for the Yield Line analysis using a value of 0.17 for β_a'' is given in Figure 30. Furthermore, Table 5 shows the likelihood function values across all increments of the proportionality constant β_3 , for each combination of Awareness and Novelty Persistence distributions. The shape parameters that maximized the likelihood function were:

$$\beta_a' = 0.27$$

$$\beta_2' = 1.00$$

$$\mu_a = 5.4 \text{ weeks}$$

$$\mu_n = 11.0 \text{ weeks}$$

$$\beta_3 = 0.30$$

The associated Log-Likelihood value for this combination of parameters was -20.324. The parameter values are similar to those of the original results, with stability reached by about 25 weeks (as before).

The distribution forms for $f_n(t_n)$ and $f_a(t_a)$ that maximized the likelihood function for the Yield Line data were a Normal distribution for the Awareness Effect and a Normal

FIGURE 29: Percent reduction in violation rate by week at Location 5 (ramp rehabilitation and wider shoulder) using the unified solution specifications of 0.30 for β_3 and a Normal distribution for $f_n(t_n)$.

FIGURE 30: Percent reduction in violation rate by week at Location 5 (Yield Line) using the updated value of $\beta_a'' = 0.17$.

$f_a(t_a) : \text{Normal}$							$f_n(t_n) : \text{Normal}$						
μ_n	β'_2	μ_a	β_3	β'_a	Log Likelihood	Rank	μ_n	β'_2	μ_a	β_3	β'_a	Log Likelihood	Rank
11.0	1.00	5.5	0.30	0.26	-20.876	1	11.0	1.00	5.5	0.30	0.26	-21.137	1
11.0	1.00	5.5	0.35	0.26	-21.597	2	11.0	1.00	5.5	0.35	0.26	-22.093	2
11.5	1.00	5.0	0.40	0.26	-22.191	3	11.5	1.00	5.5	0.40	0.26	-22.859	3
12.0	1.00	5.0	0.45	0.26	-22.579	4	11.5	1.00	5.5	0.45	0.26	-23.487	4
15.0	1.00	4.5	0.50	0.22	-22.796	5	15.0	1.00	5.0	0.50	0.22	-23.900	5
15.0	1.00	4.5	0.55	0.22	-22.898	6	15.0	1.00	5.0	0.55	0.22	-24.146	6
15.0	1.00	4.5	0.60	0.22	-23.025	7	15.0	1.00	5.0	0.60	0.22	-24.376	7
15.0	1.00	4.0	0.65	0.22	-23.132	8	15.0	1.00	4.5	0.65	0.22	-24.537	8
15.0	1.00	4.0	0.70	0.22	-23.201	9	15.0	1.00	4.0	0.70	0.22	-24.689	9
15.0	1.00	4.0	0.75	0.22	-23.293	10	15.0	1.00	4.0	0.75	0.22	-24.791	10
15.0	1.00	4.0	0.80	0.22	-23.402	11	15.0	1.00	4.0	0.80	0.22	-24.889	11
15.0	1.00	3.5	0.85	0.22	-23.441	12	15.0	1.00	4.0	0.85	0.22	-24.988	12
15.0	1.00	3.5	0.90	0.22	-23.491	13	15.0	1.00	4.0	0.90	0.22	-25.137	13
15.0	1.00	3.5	0.95	0.22	-23.571	14	15.0	1.00	4.0	0.95	0.22	-25.272	14
11.0	1.00	5.4	0.30	0.27	-20.677	Optimal	10.7	1.00	5.6	0.30	0.28	-21.050	Optimal

$f_a(t_a) : \text{Normal}$							$f_n(t_n) : \text{Uniform}$						
μ_n	β'_2	μ_a	β_3	β'_a	Log Likelihood	Rank	μ_n	β'_2	μ_a	β_3	β'_a	Log Likelihood	Rank
15.0	1.00	5.0	0.30	0.22	-23.659	10	15.0	1.00	5.5	0.30	0.22	-23.995	1
15.0	1.00	5.0	0.35	0.22	-23.693	11	15.0	1.00	5.5	0.35	0.22	-24.178	2
15.0	1.00	5.0	0.40	0.22	-23.717	13	15.0	1.00	5.5	0.40	0.22	-24.441	3
15.0	1.00	5.0	0.45	0.22	-23.730	14	15.0	1.00	5.5	0.45	0.22	-24.649	4
15.0	1.00	4.5	0.50	0.22	-23.708	12	15.0	1.00	5.5	0.50	0.22	-24.811	5
15.0	1.00	4.5	0.55	0.22	-23.610	9	15.0	1.00	5.0	0.55	0.22	-24.890	9
15.0	1.00	4.5	0.60	0.22	-23.537	8	15.0	1.00	5.0	0.60	0.22	-24.935	12
15.0	1.00	4.5	0.65	0.22	-23.485	7	15.0	1.00	4.5	0.65	0.22	-24.958	14
15.0	1.00	4.0	0.70	0.22	-23.399	6	15.0	1.00	4.5	0.70	0.22	-24.942	13
15.0	1.00	4.0	0.75	0.22	-23.316	5	15.0	1.00	4.5	0.75	0.22	-24.928	11
15.0	1.00	4.0	0.80	0.22	-23.257	4	15.0	1.00	4.0	0.80	0.22	-24.901	10
14.5	1.00	4.0	0.85	0.26	-23.198	1	15.0	1.00	4.0	0.85	0.22	-24.868	7
15.0	1.00	4.0	0.90	0.22	-23.202	2	15.0	1.00	4.0	0.90	0.22	-24.841	6
15.0	1.00	4.0	0.95	0.22	-23.227	3	15.0	1.00	4.5	0.95	0.26	-24.889	8
15.0	1.00	4.0	0.89	0.24	-22.960	Optimal	15.5	1.00	5.4	0.30	0.23	-23.673	Optimal

TABLE 5 (Part 1 of 2): Optimal parameters and associated Log-Likelihoods for each combination of β_3 , $f_a(t_a)$, and $f_n(t_n)$ at Location 5 (for the driver response curves for the Yield Line) using the updated value of $\beta''_a = 0.17$. Higher-resolution overall optimal values are given in the last row of each table.

$f_a(t_a) : \text{Normal}$				$f_n(t_n) : \text{Geometric}$			$f_a(t_a) : \text{Log-Normal}$				$f_n(t_n) : \text{Geometric}$		
μ_n	β'_2	μ_a	β_3	β'_a	Log Likelihood	Rank	μ_n	β'_2	μ_a	β_3	β'_a	Log Likelihood	Rank
0.07	1.00	5.5	0.30	0.26	-25.060	7	0.07	1.00	5.5	0.30	0.26	-25.335	1
0.07	1.00	5.5	0.35	0.26	-25.172	10	0.07	1.00	5.5	0.35	0.26	-25.641	2
0.07	1.00	5.0	0.40	0.26	-25.225	12	0.07	1.00	5.5	0.40	0.26	-25.933	3
0.07	1.00	5.0	0.45	0.26	-25.232	13	0.07	1.00	5.5	0.45	0.26	-26.174	4
0.07	1.00	5.0	0.50	0.26	-25.234	14	0.07	1.00	5.5	0.50	0.26	-26.373	5
0.07	1.00	4.5	0.55	0.26	-25.221	11	0.07	1.00	5.5	0.55	0.26	-26.535	6
0.07	1.00	4.5	0.60	0.26	-25.134	9	0.07	1.00	5.5	0.60	0.26	-26.666	7
0.07	1.00	4.5	0.65	0.26	-25.072	8	0.07	1.00	5.0	0.65	0.26	-26.728	8
0.07	1.00	4.5	0.70	0.26	-25.034	6	0.07	1.00	5.0	0.70	0.26	-26.776	9
0.07	1.00	4.0	0.75	0.26	-25.003	5	0.07	1.00	5.0	0.75	0.26	-26.814	10
0.07	1.00	4.0	0.80	0.26	-24.931	3	0.07	1.00	4.5	0.80	0.26	-26.840	11
0.07	1.00	4.0	0.85	0.26	-24.884	1	0.07	1.00	4.5	0.85	0.26	-26.843	12
0.07	1.00	4.0	0.90	0.26	-24.898	2	0.07	1.00	4.5	0.90	0.26	-26.859	13
0.07	1.00	4.0	0.95	0.26	-24.960	4	0.07	1.00	4.5	0.95	0.26	-26.895	14
0.07	1.00	3.9	0.88	0.25	-24.825	Optimal	0.07	1.00	5.4	0.30	0.24	-25.222	Optimal

TABLE 5 (Part 2 of 2): Optimal parameters and associated Log-Likelihoods for each combination of β_3 , $f_a(t_a)$, and $f_n(t_n)$ at Location 5 (for the driver response curves for the Yield Line) using the updated value of $\beta''_a = 0.17$. Higher-resolution overall optimal values are given in the last row of each table.

distribution for the Novelty Persistence Effect, just as we found in our first analysis of the data. Similarly, a Log-Normal Distribution tended to produce lower likelihood values than the Normal distribution for the Awareness effect (see Table 2), reinforcing the finding from the Larger Sign analysis that the Awareness effect is better modeled by a Normal distribution.

Using the revised results for the Yield Line data given in Table 5, we obtain a new set of combined likelihood values for both compliance strategies. These new unified results are given in Table 6, which compares the combined likelihood values associated with different values for β_3 and distributions for $f_n(t_n)$. As we did in Table 4, we only consider a Normal distribution for $f_a(t_a)$ since the Log-Normal distribution generally resulted in lower likelihood values in the analyses of both the Yield Line and the Larger Sign. For this same reason, we also exclude the Geometric distribution as a possibility for $f_n(t_n)$.

From the new unified results of Table 6, we identify the same three candidate optima that were present in Table 4. As before, a maximum likelihood is achieved at $\beta_3 = 0.83$ with a Uniform distribution for $f_n(t_n)$. Also as before, however, this optima results in μ_n hitting its upper bound for the Yield Sign curve parameters, making it a less reasonable choice than the next-ranked optima at $\beta_3 = 0.30$ with a Normal distribution for $f_n(t_n)$. Because this agrees with the initial unified results that we started with, we do not need to perform any additional iterations. Thus, we have successfully combined the results from our Yield Line and Larger Sign data to arrive at a common, maximum likelihood estimate for the value of β_3 and the distribution families of $f_n(t_n)$ and $f_a(t_a)$.

Confidence Intervals for MOE

Using Likelihood Ratio Tests to constrain β'_a to different values, we can use the corresponding p-values to identify 95% confidence intervals for β'_a in the case of the larger Yield sign and the Yield Line. For each constrained value of β'_a , we searched for a combination of μ_n , μ_a , and β'_2 that maximized the likelihood function. Based on our findings from the “Unified Results” section earlier, we used a proportionality constant of $\beta_3 = 0.30$, a Normal distribution for $f_n(t_n)$, and a Normal distribution for $f_a(t_a)$. Table 7 shows our likelihood-maximizing parameters and their associated p-values for the Larger Sign data, with Table 8 showing the same results for the Yield Line data.

The upper and lower bounds for the confidence intervals were determined by gradually stepping β'_a farther from its original (optimal) value, and identifying the first value for β'_a in each direction that resulted in a p-value lower than 2.5%. For the Larger Sign data, this resulted in a 95% confidence interval for β'_a of 0.10 to 0.26. For the Yield Line, our 95% confidence interval for β'_a was between 0.17 and 0.34. We note in both cases that the p-values decrease more rapidly above the optimal values for β'_a than below them, which is a consequence of the fact that most data points lie below β'_a (see Figure 28 and Figure 30). To shrink our confidence intervals, we would need to add more data points to the far right portions of Figures 28 and 30, in the weeks after the Driver Response Curves have stabilized.

Assuming $\beta_3 = 0.82$ with a Uniform distribution for $f_n(t_n)$

With our final results and associated confidence intervals for the Yield Line and Larger Sign analyses now complete, we take a brief look at the effect of using option (c) instead of option (a) in the “Unified Results” section. For this analysis, we will repeat the procedure described in the section on “Revisiting the Ramp Rehabilitation and Yield Line Analyses,” using solution (c) from the “Unified Results” section instead of solution (a).

$f_a(t_a) : \text{Normal}$					$f_n(t_n) : \text{Normal}$				
β_3	Log Likelihood (Larger Sign)	Log Likelihood (Yield Line)	Combined Log Likelihood	Rank	β_3	Log Likelihood (Larger Sign)	Log Likelihood (Yield Line)	Combined Log Likelihood	Rank
0.30	-17.875	-20.876	-38.751	1	0.30	-17.875	-23.659	-41.534	14
0.35	-17.455	-21.597	-39.052	6	0.35	-17.455	-23.693	-41.148	13
0.40	-17.098	-22.191	-39.290	9	0.40	-17.098	-23.717	-40.815	12
0.45	-16.829	-22.579	-39.407	11	0.45	-16.829	-23.730	-40.558	11
0.50	-16.635	-22.796	-39.431	12	0.50	-16.635	-23.708	-40.343	10
0.55	-16.574	-22.898	-39.472	13	0.55	-16.677	-23.610	-40.287	9
0.60	-16.317	-23.025	-39.342	10	0.60	-16.332	-23.537	-39.868	8
0.65	-15.929	-23.132	-39.060	7	0.65	-15.917	-23.485	-39.402	7
0.70	-15.782	-23.201	-38.983	3	0.70	-15.719	-23.399	-39.119	6
0.75	-15.745	-23.293	-39.038	5	0.75	-15.608	-23.316	-38.924	5
0.80	-15.570	-23.402	-38.972	2	0.80	-15.425	-23.257	-38.682	3
0.85	-15.555	-23.441	-38.996	4	0.85	-15.388	-23.198	-38.586	1
0.90	-15.773	-23.491	-39.265	8	0.90	-15.465	-23.202	-38.667	2
0.95	-15.997	-23.571	-39.568	14	0.95	-15.608	-23.227	-38.835	4
<i>High Resolution Analysis Around $\beta_3 = 0.30$</i>					<i>High Resolution Analysis Around $\beta_3 = 0.85$</i>				
0.25	-17.799	-20.001	-37.799	1	0.8	-15.110	-23.045	-38.155	6
0.26	-17.771	-20.149	-37.920	2	0.81	-15.101	-23.032	-38.133	4
0.27	-17.725	-20.276	-38.001	3	0.82	-15.095	-23.020	-38.115	2
0.28	-17.677	-20.411	-38.088	4	0.83	-15.093	-23.008	-38.101	1
0.29	-17.628	-20.521	-38.149	5	0.84	-15.125	-22.998	-38.122	3
0.30	-17.577	-20.677	-38.253	6	0.85	-15.162	-22.986	-38.148	5
0.31	-17.524	-20.851	-38.375	7	0.86	-15.202	-22.974	-38.176	7
0.32	-17.471	-21.027	-38.498	8	0.87	-15.244	-22.963	-38.207	8
0.33	-17.417	-21.195	-38.612	9	0.88	-15.288	-22.961	-38.249	9
0.34	-17.362	-21.354	-38.716	10	0.89	-15.330	-22.960	-38.290	10
0.35	-17.308	-21.506	-38.813	11	0.9	-15.372	-22.968	-38.340	11

TABLE 6: New Combined Log-Likelihoods from the Driver Response Curves of the Larger Sign and Yield Line for each value of β_3 and different forms of $f_n(t_n)$, after starting with the assumptions for the ramp rehabilitation analysis that $\beta_3 = 0.3$ and $f_n(t_n)$ follows a Normal distribution.

$f_a(t_a) : \text{Normal}$					$f_n(t_n) : \text{Normal}$		
μ_n	β'_2	μ_a	β_3	β'_a	Log Likelihood	D-Statistic	P-Value
12.5	1.00	2.1	0.30	0.00	-25.398	15.642	0.01%
12.5	1.00	2.1	0.30	0.01	-24.544	13.935	0.02%
12.5	1.00	2.2	0.30	0.02	-23.682	12.212	0.05%
12.5	0.96	2.2	0.30	0.03	-23.153	11.152	0.08%
12.5	0.92	2.2	0.30	0.04	-22.654	10.155	0.14%
12.5	0.84	2.1	0.30	0.05	-22.183	9.212	0.24%
12.5	0.79	2.1	0.30	0.06	-21.703	8.254	0.41%
12.5	0.74	2.1	0.30	0.07	-21.267	7.380	0.66%
12.5	0.68	2.1	0.30	0.08	-20.859	6.564	1.04%
12.5	0.64	2.1	0.30	0.09	-20.502	5.850	1.56%
12.5	0.60	2.1	0.30	0.10	-20.196	5.238	2.21%
12.5	0.56	2.1	0.30	0.11	-19.930	4.707	3.00%
12.5	0.52	2.1	0.30	0.12	-19.705	4.257	3.91%
12.5	0.48	2.1	0.30	0.13	-19.518	3.884	4.88%
12.5	0.44	2.1	0.30	0.14	-19.367	3.580	5.85%
12.5	0.40	2.1	0.30	0.15	-19.245	3.336	6.78%
12.5	0.35	2.1	0.30	0.16	-19.136	3.119	7.74%
12.5	0.25	1.6	0.30	0.17	-19.003	2.853	9.12%
12.5	0.18	1.6	0.30	0.18	-18.754	2.355	12.49%
12.5	0.13	1.6	0.30	0.19	-18.444	1.735	18.78%
12.5	0.07	1.6	0.30	0.20	-18.061	0.969	32.48%
0.1	0.00	1.6	0.30	0.21	-17.609	0.065	79.87%
0.1	0.00	1.6	0.30	0.22	-17.577	0.000	100.00%
0.1	0.00	1.6	0.30	0.23	-17.915	0.676	41.08%
0.7	0.04	2.1	0.30	0.24	-18.548	1.943	16.34%
0.1	0.00	2.1	0.30	0.25	-19.567	3.981	4.60%
0.8	0.05	2.2	0.30	0.26	-21.569	7.985	0.47%
0.1	0.00	2.2	0.30	0.27	-23.820	12.487	0.04%
0.1	0.00	2.2	0.30	0.28	-26.683	18.213	0.00%
2.5	1.00	5.4	0.30	0.29	-29.339	23.524	0.00%
2.4	1.00	5.4	0.30	0.30	-31.243	27.333	0.00%
2.1	1.00	5.4	0.30	0.31	-33.400	31.647	0.00%
2.1	1.00	5.5	0.30	0.32	-35.811	36.468	0.00%
2.0	1.00	5.5	0.30	0.33	-38.427	41.700	0.00%
1.8	1.00	5.5	0.30	0.34	-41.304	47.455	0.00%
1.7	1.00	5.5	0.30	0.35	-44.454	53.754	0.00%
1.5	1.00	5.5	0.30	0.36	-47.876	60.599	0.00%
1.4	0.95	5.5	0.30	0.37	-51.616	68.078	0.00%
1.2	1.00	5.5	0.30	0.38	-55.560	75.966	0.00%
1.0	1.00	5.5	0.30	0.39	-59.823	84.494	0.00%
0.9	0.90	5.5	0.30	0.40	-64.440	93.727	0.00%

TABLE 7: Likelihood Ratio Test p-values for constraining β'_a to different values for the Larger Sign data and the unified solution specifications of 0.30 for β_3 and a Normal distribution for $f_n(t_n)$. Dashed lines indicate the limits of the 95% Confidence Interval.

$f_a(t_a) : \text{Normal}$					$f_n(t_n) : \text{Normal}$		
μ_n	β'_2	μ_a	β_3	β'_a	Log Likelihood	D-Statistic	P-Value
15.5	1.00	4.9	0.30	0.00	-38.214	35.074	0.00%
15.5	1.00	4.9	0.30	0.01	-37.119	32.885	0.00%
15.5	1.00	4.9	0.30	0.02	-36.023	30.693	0.00%
15.5	1.00	4.9	0.30	0.03	-34.928	28.502	0.00%
15.5	1.00	4.9	0.30	0.04	-33.833	26.313	0.00%
15.5	1.00	4.9	0.30	0.05	-32.742	24.130	0.00%
15.5	1.00	4.9	0.30	0.06	-31.654	21.955	0.00%
15.5	1.00	4.9	0.30	0.07	-30.572	19.790	0.00%
15.5	1.00	4.8	0.30	0.08	-29.497	17.640	0.00%
15.5	1.00	4.8	0.30	0.09	-28.430	15.506	0.01%
15.5	1.00	4.8	0.30	0.10	-27.374	13.394	0.03%
15.5	1.00	4.9	0.30	0.11	-26.467	11.581	0.07%
15.5	1.00	4.9	0.30	0.12	-25.749	10.145	0.14%
15.5	1.00	4.9	0.30	0.13	-25.133	8.913	0.28%
15.5	1.00	4.9	0.30	0.14	-24.623	7.893	0.50%
15.5	1.00	5.0	0.30	0.15	-24.115	6.876	0.87%
15.5	1.00	5.0	0.30	0.16	-23.674	5.994	1.44%
15.5	1.00	5.0	0.30	0.17	-23.280	5.206	2.25%
15.5	1.00	5.1	0.30	0.18	-22.906	4.458	3.47%
13.6	1.00	5.1	0.30	0.19	-22.594	3.835	5.02%
13.1	1.00	5.1	0.30	0.20	-22.243	3.133	7.67%
12.6	1.00	5.2	0.30	0.21	-21.883	2.413	12.04%
12.3	1.00	5.2	0.30	0.22	-21.585	1.816	17.78%
12.0	1.00	5.2	0.30	0.23	-21.327	1.301	25.40%
11.7	1.00	5.3	0.30	0.24	-21.073	0.793	37.31%
11.5	1.00	5.3	0.30	0.25	-20.900	0.447	50.38%
11.3	1.00	5.3	0.30	0.26	-20.780	0.206	65.00%
11.0	1.00	5.4	0.30	0.27	-20.677	0.000	100.00%
10.7	1.00	5.5	0.30	0.28	-20.760	0.166	68.36%
10.5	1.00	5.5	0.30	0.29	-21.067	0.780	37.72%
10.2	1.00	5.6	0.30	0.30	-21.436	1.518	21.78%
10.0	1.00	5.7	0.30	0.31	-21.858	2.363	12.42%
9.7	1.00	5.8	0.30	0.32	-22.364	3.375	6.62%
9.5	1.00	5.9	0.30	0.33	-22.934	4.515	3.36%
9.3	1.00	5.9	0.30	0.34	-23.705	6.057	1.38%
9.1	1.00	6.0	0.30	0.35	-24.606	7.858	0.51%
8.9	1.00	6.1	0.30	0.36	-25.696	10.039	0.15%
3.1	1.00	5.5	0.30	0.37	-26.282	11.210	0.08%
3.0	1.00	5.5	0.30	0.38	-26.581	11.808	0.06%
2.9	1.00	5.6	0.30	0.39	-26.957	12.561	0.04%
2.7	1.00	5.6	0.30	0.40	-27.393	13.432	0.02%

TABLE 8: Likelihood Ratio Test p-values for constraining β'_a to different values for the Yield Line data and the unified solution specifications of 0.30 for β_3 and a Normal distribution for $f_n(t_n)$. Dashed lines indicate the limits of the 95% Confidence Interval.

From a revised analysis of the Ramp Rehabilitation data using a given value of 0.82 for β_3 and a Uniform distribution for $f_n(t_n)$, we obtain the following likelihood-maximizing shape parameters:

$$\beta_a'' = 0.16$$

$$\beta_2' = 1.00$$

$$\mu_a = 0.00 \text{ weeks (accurate to } \pm 0.05)$$

$$\mu_n = 2.25 \text{ weeks (accurate to } \pm 0.05)$$

The associated Log-Likelihood value for this combination of parameters was -17.011. The parameter values are very similar to those of the original results, which is reasonable given that the only initial condition that changed was the adjustment of β_3 from 0.83 to 0.82. As before, we note that stability is achieved after the first four weeks. Consequently, for the Yield Line analysis, we do not need to make any changes regarding which weeks of data are excluded and included for our control ramp at Location 1, since we were already excluding the first four weeks following its rehabilitation in early 2011.

Because our optimal value for β_a'' is no different from what it was when we first analyzed the Yield Line data, the results from Table 3 remain valid. Similarly, the results from Table 4 remain valid as well, since there are no updates to any of the likelihood values for either compliance strategy. This means that the original optimal values identified in the “Unified Results” section earlier are still valid. This confirms the initial unified results of our iterative solution process here, meaning that we do not need to proceed through any additional cycles. The parameter values for the Yield Line analysis associated with solution (c) of the “Unified Results” section are:

$$\beta_a' = 0.23$$

$$\beta_2' = 1.00$$

$$\mu_a = 4.1 \text{ weeks}$$

$$\mu_n = 15 \text{ weeks}$$

The corresponding Driver Response Curve defined by these parameters is shown in Figure 31.

Finally, we can use these results for the Yield Line data to establish 95% confidence intervals on the value of β_a' for the Yield Line, just as we did in the preceding section. For this, we use a proportionality constant of $\beta_3 = 0.82$, a Uniform distribution for $f_n(t_n)$, and a Normal distribution for $f_a(t_a)$. Table 9 shows our likelihood-maximizing parameters and their associated p-values for the Larger Sign data, and Table 10 shows them for the Yield Line data. The 95% confidence interval for β_a' went from 0.11 to 0.26 in the case of the Larger Sign, and from 0.14 to 0.33 in the case of the Yield Line.

Note that both of the MOE confidence intervals are very similar to those obtained when we selected solution (a) instead of solution (c) in the “Unified Results” section, suggesting that our MOE is not highly sensitive to the value of β_3 , nor is it very sensitive to whether we specify $f_n(t_n)$ as a Uniform distribution or Normal distribution.

FIGURE 31: Percent reduction in violation rate by week at Location 5 (Yield Line)
 using $\beta_a'' = 0.16$, $\beta_3 = 0.82$ with a Uniform distribution for $f_n(t_n)$

$f_a(t_a) : \text{Normal}$					$f_n(t_n) : \text{Uniform}$		
μ_n	β'_2	μ_a	β_3	β'_a	Log Likelihood	D-Statistic	P-Value
12.5	1.00	1.8	0.82	0.00	-28.036	25.970	0.00%
12.5	1.00	1.9	0.82	0.01	-26.480	22.859	0.00%
12.5	1.00	2.0	0.82	0.02	-25.011	19.920	0.00%
12.5	1.00	2.1	0.82	0.03	-23.628	17.155	0.00%
12.5	1.00	2.2	0.82	0.04	-22.335	14.567	0.01%
12.5	1.00	2.3	0.82	0.05	-21.115	12.128	0.05%
12.5	1.00	2.4	0.82	0.06	-20.081	10.059	0.15%
12.5	1.00	2.6	0.82	0.07	-19.337	8.572	0.34%
12.5	1.00	2.6	0.82	0.08	-18.785	7.467	0.63%
11.1	1.00	2.7	0.82	0.09	-18.456	6.811	0.91%
10.2	1.00	2.9	0.82	0.10	-18.030	5.959	1.46%
<hr/>							
9.4	1.00	3.0	0.82	0.11	-17.604	5.107	2.38%
8.7	1.00	3.2	0.82	0.12	-17.213	4.325	3.76%
8.1	1.00	3.3	0.82	0.13	-16.827	3.552	5.95%
7.6	1.00	3.4	0.82	0.14	-16.491	2.880	8.97%
7.1	1.00	3.5	0.82	0.15	-16.168	2.234	13.50%
6.5	1.00	3.6	0.82	0.16	-15.848	1.594	20.68%
5.9	1.00	3.7	0.82	0.17	-15.526	0.949	33.00%
5.5	1.00	3.7	0.82	0.18	-15.241	0.381	53.72%
4.9	1.00	3.7	0.82	0.19	-15.051	0.000	100.00%
4.3	1.00	3.7	0.82	0.20	-15.084	0.065	79.86%
4.2	1.00	3.8	0.82	0.21	-15.328	0.553	45.70%
3.9	1.00	3.9	0.82	0.22	-15.713	1.324	25.00%
3.1	1.00	3.9	0.82	0.23	-16.057	2.011	15.61%
2.8	1.00	3.9	0.82	0.24	-16.448	2.795	9.46%
2.7	1.00	4.0	0.82	0.25	-17.163	4.224	3.99%
2.4	1.00	4.0	0.82	0.26	-18.267	6.433	1.12%
<hr/>							
2.3	0.95	4.0	0.82	0.27	-19.513	8.925	0.28%
2.2	0.85	4.0	0.82	0.28	-20.936	11.771	0.06%
2.2	0.79	4.0	0.82	0.29	-22.542	14.981	0.01%
2.2	0.73	4.0	0.82	0.30	-24.373	18.645	0.00%
2.2	0.60	4.0	0.82	0.31	-26.423	22.744	0.00%
1.8	0.47	3.9	0.82	0.32	-28.635	27.168	0.00%
1.7	0.40	4.0	0.82	0.33	-31.060	32.019	0.00%
1.7	0.23	4.0	0.82	0.34	-33.692	37.283	0.00%
4.8	1.00	9.9	0.82	0.35	-35.063	40.023	0.00%
4.8	1.00	10.4	0.82	0.36	-36.322	42.542	0.00%
4.8	1.00	11.1	0.82	0.37	-37.601	45.100	0.00%
5.0	1.00	12.5	0.82	0.38	-38.752	47.402	0.00%
4.8	1.00	12.5	0.82	0.39	-39.945	49.788	0.00%
4.7	1.00	12.5	0.82	0.40	-41.273	52.444	0.00%

TABLE 9: Likelihood Ratio Test p-values for constraining β'_a to different values for the Larger Sign data and the alternate unified solution specifications of 0.82 for β_3 and a Uniform distribution for $f_n(t_n)$. Dashed lines indicate the limits of the 95% Confidence Interval.

$f_a(t_a) : \text{Normal}$					$f_n(t_n) : \text{Uniform}$		
μ_n	β'_2	μ_a	β_3	β'_a	Log Likelihood	D-Statistic	P-Value
15.5	1.00	3.7	0.82	0.00	-39.642	34.593	0.00%
15.5	1.00	3.7	0.82	0.01	-38.547	32.402	0.00%
15.5	1.00	3.7	0.82	0.02	-37.451	30.211	0.00%
15.5	1.00	3.7	0.82	0.03	-36.356	28.020	0.00%
15.5	1.00	3.7	0.82	0.04	-35.262	25.832	0.00%
15.5	1.00	3.7	0.82	0.05	-34.171	23.651	0.00%
15.5	1.00	3.7	0.82	0.06	-33.085	21.478	0.00%
15.5	1.00	3.7	0.82	0.07	-32.005	19.318	0.00%
15.5	1.00	3.7	0.82	0.08	-30.932	17.173	0.00%
15.5	1.00	3.6	0.82	0.09	-29.908	15.125	0.01%
15.5	1.00	3.6	0.82	0.10	-28.979	13.266	0.03%
15.5	1.00	3.6	0.82	0.11	-28.067	11.442	0.07%
15.5	1.00	3.6	0.82	0.12	-27.175	9.659	0.19%
15.5	1.00	3.6	0.82	0.13	-26.306	7.922	0.49%
15.5	1.00	3.6	0.82	0.14	-25.464	6.236	1.25%
15.5	1.00	3.6	0.82	0.15	-24.701	4.711	3.00%
15.5	1.00	3.6	0.82	0.16	-24.132	3.573	5.87%
15.5	1.00	3.6	0.82	0.17	-23.633	2.575	10.86%
15.5	1.00	3.8	0.82	0.18	-23.219	1.747	18.63%
15.5	1.00	3.8	0.82	0.19	-22.917	1.142	28.52%
15.5	1.00	3.9	0.82	0.20	-22.675	0.658	41.72%
15.5	1.00	3.9	0.82	0.21	-22.493	0.295	58.72%
15.5	1.00	4.0	0.82	0.22	-22.382	0.072	78.89%
15.5	1.00	4.1	0.82	0.23	-22.346	0.000	100.00%
15.5	1.00	4.1	0.82	0.24	-22.427	0.163	68.67%
14.5	1.00	4.2	0.82	0.25	-22.707	0.722	39.56%
13.7	1.00	4.2	0.82	0.26	-23.026	1.361	24.33%
10.7	1.00	4.3	0.82	0.27	-23.327	1.963	16.12%
9.8	1.00	4.3	0.82	0.28	-23.516	2.340	12.61%
9.3	1.00	4.4	0.82	0.29	-23.733	2.775	9.57%
8.9	1.00	4.4	0.82	0.30	-23.977	3.262	7.09%
8.7	1.00	4.4	0.82	0.31	-24.248	3.804	5.11%
8.3	1.00	4.5	0.82	0.32	-24.572	4.452	3.49%
8.2	1.00	4.5	0.82	0.33	-25.061	5.431	1.98%
7.3	1.00	4.6	0.82	0.34	-25.698	6.704	0.96%
7.2	1.00	4.6	0.82	0.35	-26.390	8.089	0.45%
5.7	1.00	4.6	0.82	0.36	-27.005	9.318	0.23%
4.8	1.00	4.5	0.82	0.37	-27.538	10.384	0.13%
4.4	1.00	4.5	0.82	0.38	-28.069	11.447	0.07%
2.6	0.90	4.1	0.82	0.39	-28.542	12.392	0.04%
4.8	0.09	3.4	0.82	0.40	-28.880	13.069	0.03%

TABLE 10: Likelihood Ratio Test p-values for constraining β'_a to different values for the Yield Line data and the alternate unified solution specifications of 0.82 for β_3 and a Uniform distribution for $f_n(t_n)$. Dashed lines indicate the limits of the 95% Confidence Interval.

Other Prediction Models

Making accurate predictions of driver behavior in the weeks for the “after” period is crucial to obtaining an accurate representation of the Driver Response Curve and evaluation of the compliance strategy’s effectiveness. Although we have attempted to identify models, $P(t)$, that minimize the prediction risk as evaluated through LOOCV, we recognize that our preceding model search was limited to combinations of variables that fit the form of Equation 10. We now use our data from Location 8 (for the Larger Sign analysis) to consider three alternative modeling approaches that depart from the constraints of Equation 10, and evaluate whether any of them result in better predictions of driver behavior in a given week.

Disaggregate by Hour

Our first alternative model follows the same principle as Equation 10, but uses data that has been disaggregated into hours rather than three-hour periods. As a result, we have 29 data points in our “before” data set (including disaggregate data for each potential covariate), as opposed to the 10 we originally had. One hourly data point was not available at the experimental ramp due to an equipment issue that affected about 30 minutes of video footage between 12:15 and 12:45 PM on Nov 1. In four other hours, gaps existed for one of the covariates; these were filled with the average of the remaining values for the affected variable in each case.

To identify the best prediction model using the hourly data, we again used LOOCV to try every possible combination of up to four variables, and found that the following model minimized the prediction risk:

$$P(t) = \text{Logit}^{-1}[1.468x_1(t) + 0.942x_2(t) - 0.651x_3(t) + 0.185] \quad (15)$$

where

$x_1(t)$ is the percent of time the on-ramp flow exceeded 80 vehicles per 5 minutes for hour t , using bins of 5 minutes.

$x_2(t)$ is the percent of time the freeway speed dropped below 55 mph at the on-ramp for hour t , using bins of 5 minutes.

$x_3(t)$ is the compliance rate at Location 9 in hour t , where a gap for the last hour of Nov 8 was filled with an average of the rates from all other hours.

The following table lists the properties of the model coefficients, with the raw data given in Appendix H.

Variable	Coefficient	Standard Error	Significance
Percent of time the On-Ramp Flow exceeds 80 veh/5 min	1.468	1.040	0.104
Percent of time the freeway speed is below 55 mph	0.942	1.053	0.203
Compliance rate at Location 9 Control Ramp	-0.651	0.879	0.243
Constant	0.185	0.454	0.349

This model had an average prediction risk of 0.005967 for each hour, a prediction error mean of $\mu_{cv} = -0.0006$, and a prediction error standard deviation of $\sigma_{cv} = 0.0786$. For comparison, a constant-only model using hourly data had an average risk of 0.005974, a prediction error mean of $\mu_{cv} = 0.000$, and a prediction error standard deviation of $\sigma_{cv} = 0.0787$. From a paired t -test, the hypothesis that Equation 15 does no better than the constant-only model had a p -value = 0.496, implying that Equation 15 is not significantly better than a constant model despite having the lowest prediction risk out of all the hourly models we tested. The hourly predictions from Equation 15 also showed no evidence of serial correlation at a 5% significance level, with a Durbin Watson statistic of 2.761.

This average prediction risk for Equation 15 cannot be directly compared to the 0.000457 from Equation 10 because the two models are validated against different data sets (i.e., hourly data or three-hour aggregated data). However, we can use our hourly model to make compliance predictions for each hour, and from this we can construct predictions for each week. The new LOOCV process is:

1. Estimate the parameters of Equation 15 using all available hours in the “before” data except hour t' , and use this model and the covariate data for t' to predict the proper yield rate at the experimental ramp in hour t' .
2. Repeat (1) for all hours in the “before” data set.
3. Combine the predicted rates for all three hours in a particular week, using the observed total number of yield events (i.e., “proper yields” and “failures to yield”) in each hour to obtain a weighted average rate for that week.
4. Compare the weighted averages from (3) to the observed yield rates for each week, and use a squared error loss function to evaluate the risk associated with the weekly predictions from Equation 15.

When we follow this procedure, we obtain the weekly predictions for compliance rates shown in Figure 32 and can make appropriate comparisons to Equation 10. Error bars in the figure represent 90% confidence intervals. The average risk for weekly compliance rate predictions from Equation 15 was 0.000583, which indicates that it performs worse than Equation 10. This may be a consequence of the higher degree of discretization that occurs for some of the covariate data when it is disaggregated into hours—specifically, the variables that are measured over 5-minute periods. From a paired t -test, the hypothesis that Equation 10 does no better than Equation 15 has a p -value of 0.201, suggesting that Equation 10 is the better choice of the two.

Arrival Situation Model

As a departure from the models we have tried so far, we now attempt to model driver compliance based on the particular scenario each red-path driver was presented with prior to either yielding properly or failing to yield. The conditions leading up to either of these events fall into one of the five following categories for each “subject vehicle” or red-path driver:

- a) Subject vehicle is following someone who yielded properly.
- b) Subject vehicle is following someone who failed to yield.
- c) Subject vehicle is following someone who did not need to do anything.
- d) Subject vehicle is not following anyone, and arrives at the yield point in the presence of cross-traffic
- e) Subject vehicle is not following anyone, and arrives at the yield point with no cross-traffic present.

FIGURE 32: Comparison of the LOOCV predictions from the Hourly Model to the LOOCV predictions from the Risk-Minimizing Aggregate Model (i.e., Equation 10) and the observed compliance rates in each week for Location 8 (for the 10 weeks in the “before” period of the Larger Sign analysis).

These five situations represent an exhaustive set of possibilities for a particular subject vehicle. Only those drivers who yielded or failed to yield are included; red-path drivers who did not make a choice between one of those two options are not categorized (e.g., red-path drivers who proceeded through the yield point when no yellow-path traffic was present). For the purposes of these definitions, “following” meant that the headway in front of the subject vehicle was four seconds or less. Additionally, cross-traffic is considered “present” if any yellow-path vehicles went past the merge point in the four seconds leading up to the subject vehicle’s arrival to the yield/merge area.

In rare circumstances, a leading vehicle (i.e., a vehicle less than four seconds ahead of a subject vehicle) would pause noticeably at the yield point even when no yellow-path traffic was present, prior to proceeding through the merge area. We expect that the psychological impact on the subject driver associated with the vehicle in front stopping briefly would be the same regardless of whether a yellow-path vehicle was actually present or not—specifically, that the subject driver in either case would be primed to stop as well. As a result, a subject vehicle in this situation was classified into group (a) rather than (c), even though the leading vehicle technically did not “yield properly.”

Table 11 shows the compliance rates for each of the “before” weeks at Location 8, broken down by category. The volume of subject vehicles associated with each week and category is also given in the figure. These volumes are used to calculate weighted averages of the yield rates for each category across all ten weeks of data. From these weighted averages, we obtain the following model that predicts compliance based on the particular scenario each subject driver experienced upon arriving at the yield point:

$$P(t) = 0.425n_a(t) + 0.540n_b(t) + 0.395n_c(t) + 0.538n_d(t) + 0.704n_e(t) \quad (16)$$

where n_a represents the fraction of drivers that fall into category (a) in week (t), and the other variables are defined analogously. The coefficients represent the overall compliance rates for drivers in each category. Thus, the drivers in category (e) were most likely to yield properly when conditions warranted it, and drivers in category (c) were most likely to fail to yield.

To evaluate the risk associated with the predictions of Equation 16, we use a LOOCV procedure as before. We begin by recalculating the coefficients with one week of data excluded, and use the resultant model to predict the compliance rate for that validation week based on the proportion of drivers that fell into each category for that week. We evaluate prediction risk as the squared difference between the predicted and observed compliance rates, repeat the procedure so that every week is treated as the “validation” week once, and compute an average risk estimate from the results of each iteration. Figure 33 compares the predicted and observed compliance rates by week, where each week is predicted from a version of Equation 16 that did not use that week’s data when estimating the coefficients of the model. Error bars in the figure represent 90% confidence intervals.

From this procedure, we found that Equation 16 had an average prediction risk of 0.000626, a prediction error mean of $\mu_{cv} = -0.000751$, and a prediction error standard deviation of $\sigma_{cv} = 0.0264$. This is lower than the risk associated with a constant-only model (given in the “Prediction Model Formulation” section of the Larger Sign analysis), but worse than the predictions given by Equation 10. From a paired t -test, the hypothesis that Equation 10 does no better than Equation 16 has a p -value of 0.106, suggesting that Equation 10 is the better choice of the two.

	NOT FOLLOWING		FOLLOWING			Overall Yield Rate
	No Cross Traffic	Cross Traffic	No action needed	Leader Failed to Yield	Leader Yielded	
Oct 18	42.3% (52)	48.5% (33)	37.5% (56)	50% (30)	100% (8)	45.8% (179)
Oct 25	46.4% (69)	62.5% (16)	32.1% (56)	62.5% (16)	88.9% (9)	47% (166)
Nov 1	40.5% (42)	52.2% (23)	44.4% (54)	60% (10)	80% (5)	47% (134)
Nov 8	42.6% (54)	63.6% (33)	45.7% (46)	55.6% (18)	45.5% (11)	49.4% (162)
Nov 15	40.3% (67)	56.7% (30)	41.1% (56)	61.5% (26)	75% (8)	47.6% (187)
Nov 22	41.1% (56)	55.6% (36)	43.4% (53)	47.8% (23)	71.4% (7)	46.9% (175)
Nov 29	36.1% (61)	60.9% (23)	29.4% (68)	61.9% (21)	85.7% (7)	41.7% (180)
Dec 6	48.2% (56)	53.6% (28)	44.9% (49)	56.3% (16)	55.6% (9)	49.4% (158)
Dec 13	40.6% (69)	48.1% (27)	38.3% (47)	56% (25)	50% (6)	43.7% (174)
Dec 20	46.9% (64)	39.1% (23)	42.9% (49)	76.2% (21)	63.6% (11)	49.4% (168)
Total	42.5% (590)	54% (272)	39.5% (534)	58.3% (206)	70.4% (81)	

TABLE 11: Compliance rates by week and arrival situation.
Number of events for each category and week are indicated in parentheses.

FIGURE 33: Comparison of the LOOCV predictions from the Arrival Situation Model to the LOOCV predictions from the Risk-Minimizing Aggregate Model (i.e., Equation 10) and the observed compliance rates in each week for Location 8 (for the 10 weeks in the “before” period of the Larger Sign analysis).

Driver Choice Model for Location 8

Taking our hourly model (i.e., Equation 15) one step further, we now disaggregate by individual driver and use a choice model approach to try to predict driver behavior from week to week. We use a Logit specification for the model with a binary dependent variable (i.e., whether the vehicle yielded properly when given the choice, or failed to yield). When all ten weeks of data are considered, we have 1682 observations at Location 8 to draw from for our model. To develop our choice model, we use the following approach:

1. Using all of the observations, we use SPSS to identify the combination of predictor variables and the associated variable coefficients that maximize the Log-Likelihood function for each driver's observed choice.
2. We obtain model predictions for yield rates in a given week by considering the observations from that week and summing the probabilities of each of those drivers yielding properly (dividing this sum by the total number of subject drivers for that week results in a compliance rate estimate).
3. We estimate the risk associated with the best choice model from Step 1 through LOOCV. For this, we remove the data for a particular week from the data set and then find the optimal coefficients for the choice model using only the remaining nine weeks of data. We then use this model to predict each driver's behavior from the excluded week, combine the probabilities to obtain an overall yield rate estimate for that week (see Step 2), and compare this prediction to the observed yield rate. We repeat this for each of the ten weeks of data, with our measure of risk being the average of the squared residuals between the predicted and observed rates for each "left out" week.
4. We attempt to find a new version of the optimal model from Step 1 that has a lower risk associated with it. For this, we try removing variables that are not significant on a 10% level one at a time from the model, find the model with the maximum Log-Likelihood among all the models that have one covariate removed, and repeat the LOOCV risk estimate for this new, smaller model (i.e., Step 3).
5. We repeat Step 4 for combinations of two variables at a time instead of just one, and for three variables at a time, etc., until we have tested removing all possible combinations of those variables that were not significant at a 10% level in the original model.
6. We compare the LOOCV risk estimates among all of the models tested, and select the one with the lowest risk as our best model. Our goal is to find a model that has a lower risk than that of the aggregate model that we found earlier (i.e., Equation 10).

Logit choice probabilities. The probability of a particular driver yielding properly is given by:

$$P(\text{Yielding} = 1) = \frac{e^{V(1)}}{e^0 + e^{V(1)}} \quad (17)$$

$$V(1) = \text{Linear function of the predictor variables} \quad (18)$$

The probability of a driver failing to yield is therefore:

$$P(\text{Yielding} = 0) = 1 - P(\text{Yielding} = 1) = \frac{e^0}{e^0 + e^{V(1)}}$$

Covariates. In developing our choice model, we make slight modifications to our original set of potential predictor variables to accommodate vehicle-level disaggregation. As we did for Equations 10 and 15, we do not allow more than one variable in a particular category (i.e., freeway speed, on-ramp flow) to enter the model at the same time. The available covariates for our choice model are:

- a) On-ramp flow for the five-minute window that the observation occurred in.
- b) The natural logarithm of the on-ramp flow for the five-minute window that the observation occurred in.
- c) Dummy variables indicating whether the on-ramp flow exceeded 60 vehicles, 70 vehicles, and 80 vehicles for the five-minute window that the observation occurred in.
- d) The observed mainline freeway speed past the on-ramp for the five-minute window that the observation occurred in.
- e) Dummy variable indicating whether the mainline freeway speed fell below 55 mph for the five-minute window that the observation occurred in.
- f) Dummy variable indicating whether it was sunny in the middle of the five-minute window that the observation occurred in.
- g) Observed yield rate at a nearby control ramp on the same side of the freeway for the hour that the observation occurred in.
- h) Observed yield rate at a nearby control ramp at the same interchange for the hour that the observation occurred in.

Optimal Choice Model. After evaluating 72 different choice models, we identified the following as our optimal utility function (i.e., our optimal version of Equation 18, based on which model minimized the prediction risk for the compliance rates each week):

$$V(1) = -0.012x_1 + 0.125x_2 + 1.503x_3$$

where

x_1 is the mainline freeway speed past the on-ramp for the five-minute window that the observation occurred in.

x_2 is variable (c) from the earlier list of covariates, with a 60-vehicle threshold.

x_3 is the compliance rate at Location 1 for the hour that the observation occurred in.

The details of our search for this optimal model, following the procedure outlined earlier, are given in Appendix I. In summary, each model was evaluated first on its ability to reliably predict individual driver choices, and then on its ability to predict weekly yield rates from individual driver behavior. The following table lists the properties of the model coefficients.

Variable	Coefficient	Standard Error	Significance
On-Ramp Flow is above 60 veh/5 min	.125	.110	.254
Freeway Speed passing on-ramp	-.012	.005	.018
Hourly yield rate at Location 1 Control Ramp	1.503	.794	.058

The coefficient for the control ramp is positive, which is as we expect (since any unobserved variables that cause a decrease in compliance at a nearby ramp could be expected to cause a decrease, not an increase, in compliance at the subject ramp as well). The freeway speed coefficient suggests that as the freeway becomes more congested, people become more likely to yield properly. This is similar to what we found in Equation 10. Finally, the on-ramp volume coefficient suggests that as the traffic using the on-ramp increases, compliance increases as well. This is contrary to what we found in Equation 11 (which was for a different Location), and may reflect the fact that higher volumes make it more likely that a yellow-path vehicle will cross the red-path vehicle's field of view before the latter arrives at the yield point, making the red-path vehicle more aware of the need to yield properly if necessary once it arrives at the Yield point.

For the LOOCV procedure we use to evaluate our optimal choice model, we begin by excluding one week's observations from the data set prior to solving for the coefficients that maximize the likelihood function. The Logit model is then applied to the observations from the excluded week to predict the behavior of individual drivers for that week, and from this we obtain a compliance rate prediction for that week. This prediction is compared to the observed proper yielding rate, with the difference representing the prediction risk. This procedure is then repeated for all 10 weeks of data. The average risk across all weeks represents the prediction risk associated with the model, where a higher risk suggests weaker predictive capability.

From LOOCV, we evaluate the risk associated with our optimal choice model as 0.0004478 with respect to weekly compliance rate predictions, making it only slightly better than Equation 10. Our choice model prediction error had a mean of $\mu_{cv} = -0.00184$ and a standard deviation of $\sigma_{cv} = 0.0222$. Using a paired two-sample t -test for the squared residuals from our optimal choice model and the risk-minimized aggregate model (i.e., Equation 10), we find no evidence that the choice model is better than the aggregate model on a 10% significance level, and obtain a p -value of 0.489 for the statistical test.

Figure 34 shows the predictions for each week from the LOOCV procedure (i.e., when that week has been "left out" of the model development, similar to the approach used to produce Figure 33), giving us greater insight into each model's ability to accurately predict the yield rate in a given week. For example, the gray bar (i.e., the leftmost bar) for Oct 18 represents the risk-minimized aggregate model prediction for that week, when the model (i.e., Equation 10) was not allowed to use the data from Oct 18 when solving for the least-squares coefficients. In Figure 34, the choice model has a better fit, with a Sum of Squared Residuals of 0.004478 compared to 0.006441 for the aggregate model (Equation 10). This reflects our earlier finding that the choice model was slightly better at making weekly compliance rate predictions than Equation 10 was, though not on a statistically significant level. Error bars in the figure represent 90% confidence intervals for the predictions.

Driver Choice Model for Location 5

Having found a choice model (i.e., Equations 17 and 18) that performs as well as the risk-minimizing aggregate model for Location 8, we next seek to find a similar choice model that performs as well as (or better than) Equation 11 for Location 5.

Since the aggregate model and the optimal choice model both performed comparably for Location 8, we need to know if the same is true for Location 5 before coming to a conclusion about which approach to use. If the aggregate model and the optimal choice model perform to about the same level for Location 5, or if the optimal choice model performs better, then our preferred model would be the choice model, as it seems a more intuitive approach. However, if

FIGURE 34: Comparison of the LOOCV predictions from the Choice Model to the LOOCV predictions from the Risk-Minimizing Aggregate Model (i.e., Equation 10) and the observed compliance rates in each week for Location 8 (for the 10 weeks in the “before” period of the Larger Sign analysis).

the choice model for Location 5 performs significantly worse than the risk-minimized aggregate model, our preferred models would remain Equations 10, 11, and 12.

We use a process analogous to the one used in the previous section (see “Driver Choice Model for Location 8”). One difference, however, is that we now have 704 observations to draw from for our Location 5 model when all nine weeks are considered. This is one week fewer than we had for Location 8, because a camera problem occurred during one week at Location 5.

Optimal Choice Model. After evaluating 72 different choice models, we identified the following as our optimal utility function for Location 5:

$$V(1) = 0.042x_1 - 2.982x_2 \quad (19)$$

where

x_1 is mainline freeway speed past the on-ramp for the five-minute window that the observation occurred in.

x_2 is the compliance rate at Location 6 for the hour that the observation occurred in.

The details of our search for this optimal model, following the procedure outlined earlier, are given in Appendix I. The following table lists the properties of the model coefficients.

Variable	Coefficient	Standard Error	Significance
Freeway Speed passing on-ramp	.042	.013	.001
Hourly yield rate at Location 6 Control Ramp	-2.982	1.128	.008

The above table raises some concerns with respect to the signs of the coefficients. First, the coefficient for the speed variable is positive in this model, whereas it was negative in the optimal choice model we found for the other experimental ramp at Location 8. This positive coefficient suggests that on-ramp traffic becomes less likely to yield as the freeway becomes congested. This may be a result of drivers becoming mentally distracted when they observe that the freeway is moving significantly slower than anticipated (perhaps because they begin to think about possible alternate routes or the consequences of the unexpected travel delay). This is similar to the findings from Equations 10 and 12, where the coefficients suggested that compliance decreased when congestion increased at Location 5, whereas the opposite relationship held for Location 8.

Another concern is the negative coefficient on the control ramp term. It is sensible to assume that the coefficient would be positive, since any unobserved variables that cause a decrease in compliance at a nearby ramp should be expected to have the same effect (i.e., a reducing effect) on compliance at the subject ramp as well. Since our optimal choice model for Location 8 had a positive coefficient in front of the control ramp variable rather than a negative one, the negative coefficient in our choice model for Location 5 is cause for concern. However, this negative coefficient is consistent with our results for Equations 11 and 12, as both equations also had negative coefficients in front of the compliance rate variables for Location 6.

From LOOCV, we evaluate the risk associated with our choice model for Location 5 (i.e., Equation 17 and 19) as 0.005028 with respect to weekly compliance rate predictions, making it an order of magnitude worse than Equation 11. Our optimal choice model prediction error had a mean of $\mu_{cv} = -0.00764$ and a standard deviation of $\sigma_{cv} = 0.07477$. Using a paired two-sample t -test for the squared residuals (with respect to weekly compliance rate predictions) for our optimal choice model and the risk-minimized aggregate model for Location 5 (i.e., Equation 11), we find that the aggregate model performs better than the choice model on a 5% significance level, and obtain a p-value of 0.0431 for the statistical test.

Figure 35 shows the predictions for each week from the LOOCV procedure (i.e., when that week has been “left out” of the model development, analogous to the approach used for Figure 34), giving us greater insight into each model’s ability to accurately predict the yield rate in a given week. Error bars in the figure represent 90% confidence intervals. Here, the aggregate risk-minimized model (i.e., Equation 11) has a better fit, with a Sum of Squared Residuals of 0.004025 compared to 0.045249 for the choice model. This reflects our earlier finding that Equation 11 (i.e., our aggregate model) was significantly better at making weekly compliance rate predictions than Equations 17 and 19 (i.e., our choice model).

Conclusions. In the case of Location 8, the choice model performed slightly better than the risk-minimized aggregate model (i.e., Equation 10) in terms of predictions of weekly compliance rates. However, since the improvement in performance was not statistically significant, either model would be an acceptable choice for Location 8.

In the case of Location 5, the choice model performed worse than the risk-minimizing aggregate model in terms of both risk and goodness of fit. A statistical analysis showed that the choice model was in fact significantly worse than the aggregate model in terms of risk (i.e., predictive capability) at the 5% level. Thus, for Location 5, the better choice would be the aggregate model (i.e., Equation 11).

Since we seek to have a consistent modeling approach between the two locations, we conclude that the better option is the aggregate model approach, as it offers significant improvements in quality of predictions for Location 5, and is no worse (on a statistically significant level) than the choice model with respect to predictions for Location 8.

Comparison to Traditional Before-and-After Approach

We now compare our MOEs for a larger Yield sign and a Yield Line to those obtained using a more traditional before-and-after approach to demonstrate the improved accuracy of our framework. To make the comparisons reasonable, we will treat our analyses of the Yield Line and the Larger Sign separately. Therefore, instead of using the results we obtained in the “Unified Results” section, we will use the individual likelihood-maximizing results for each compliance strategy as identified in each one’s respective analysis section. Furthermore, we will apply the original results for the ramp rehabilitation (i.e., $\beta_a'' = 0.16$) to the Yield Line data, since our findings in the Yield Line standalone analysis were based on this result.

Larger Signage

To reiterate, we found the larger signage to be 19% effective (according to Equation 6) in the long-term. When we disregard both the Driver Response Curve (i.e., Equation 2) and the Baseline Compliance Prediction Model (i.e., Equation 10), our revised MOE (following Equation 6) becomes 14.2%, which understates the effectiveness of the larger signage. As some

FIGURE 35: Comparison of the LOOCV predictions from the Choice Model to the LOOCV predictions from the Risk-Minimizing Aggregate Model (i.e., Equation 11) and the observed compliance rates in each week for Location 5 (for the 9 weeks in the “before” period).

researchers use a shorter “after” period than 30 weeks (4), we also examine what our MOE would be if our “after” period were truncated to 5 weeks. In this case, it becomes even less accurate at 8.7%, as a result of the trend of increasing compliance early on (caused by the Awareness effect).

Alternatively, since other literature suggests that stability is reached after 7 weeks (3), we also examine what our MOE would be if we disregarded data from the first 7 “after” weeks. In this case, it becomes more accurate at 20.4%. However, there is no assurance that this approach will work in general, as Figure 16 illustrates that conditions may not have stabilized by this time. For instance, if we include the Baseline Compliance Prediction Model in our analysis (which incorporates the best available control group, among other covariates) but continue to exclude the Driver Response Curve, our MOE excluding the first 7 weeks of data becomes 26.3%. This MOE overstates the effectiveness of the larger sign due to the presence of the novelty peak in Week 8.

If we use only the Baseline Compliance Prediction Model (without the Driver Response Curve) but extend our analysis period to include all weeks in the “after” period, we obtain an estimated MOE of 19.2%. However, this does not suggest that such an approach would be reliable for evaluating the effectiveness of a compliance strategy in general—rather, it is simply the result of the Driver Awareness and Novelty Effects balancing out, since both have similar mean durations and therefore have a tendency to cancel each other out when all weeks are combined. If, on the other hand, one of these effects had been longer than the other, it would have dominated the overall average for all weeks, as we will see in the following section that examines the Yield Line data (where the Novelty Effect dominates the Driver Awareness effect and results in an overestimate of the compliance strategy’s effectiveness). Finally, when we use the Baseline Compliance Prediction Model and restrict ourselves to only five weeks in the “after” period as some researchers have done in the literature, we obtain an MOE of 13.9%, which under-represents the compliance strategy’s effectiveness due to the Awareness Effect.

Yield Line

Through our framework, we found the Yield Line to be 26% effective per Equation 6. If we had done a simple before-and-after analysis using the compliance data for Location 5 but disregarding both the Driver Response Curve (i.e., Equation 2) and the Baseline Compliance Prediction Model (i.e., Equations 11 and 12), we would have obtained an estimated long-term effectiveness of 20.7%, which underestimates the effectiveness of the strategy (as before, in the case of the larger sign). However, although this method resulted in an understatement of the effectiveness for both the Yield Line and the Larger Sign, there is no guarantee that it would also result in an underestimate for other strategies as well.

Alternatively, when we consider an abbreviated “after” period of only five weeks, we obtain an estimated MOE of -0.8%, which significantly misrepresents the effect of the Yield Line. This low estimate is due largely to the Driver Awareness effect, which delays the observed response (and therefore any improvement) by about 5.5 weeks on average, and to the lack of a Baseline Compliance Prediction Model. For the Larger Sign data, considering just the first five weeks and disregarding the Baseline Compliance Prediction Model resulted in an underestimate of the true MOE, but by a lesser degree due to the shorter average duration for the Awareness Effect (i.e., 3.7 weeks).

If we had instead looked at only the weeks beyond the first seven in the “after” period, we would have obtained an estimated MOE of 28.2% for the Yield Line. The lack of a Baseline

Compliance Prediction Model (i.e., $P(t)$) for Location 5 has a reducing effect on the MOE estimate (as seen earlier with our MOE estimate of -0.8%), but the exclusion of the first seven weeks has a boosting effect due to the presence of the Novelty Effect in weeks 7-25 (see Figure 26), resulting in an overall MOE that is not far from the value we obtained using our framework. However, there is no guarantee that the Novelty Effect will cancel out the impact of excluding $P(t)$ as it has done here, since the character of those effects depends on the Driver Response Curve parameters and the coefficients in the $P(t)$ equation.

Finally, if we consider $P(t)$ before measuring the effectiveness of the Yield Line following a simple before-and-after approach (but continue to exclude the Driver Response Curve), we obtain the following estimates for the MOE: 33.5% when all weeks are considered, 15.2% when just the first five weeks are considered, and 39.7% when the first seven weeks are excluded. As expected based on our previous results, excluding the first seven weeks results in an overestimate of the MOE due to the pronounced Novelty Effect (see Figure 26), whereas analyzing just the first five weeks results in an underestimate due to the Awareness Effect only reaching about half of its full strength by this time (see Figure 26).

From these results, we conclude that the Baseline Compliance Prediction Model and the Driver Awareness Curve of our framework both play a significant role in obtaining reasonable MOE estimates for a given compliance strategy.

7. DISCUSSION

In this dissertation, we have examined the problems with using a traditional before-and-after approach to evaluate the effectiveness of traffic compliance strategies, and have developed a framework to overcome these issues and improve accuracy. For example, we have shown that researchers need to look carefully at disaggregate (e.g., by week) data to estimate when stability is reached for their particular compliance strategy being evaluated, as the common 4-8 week assumption has been shown here to be untenable; simply waiting eight weeks before collecting data is not sufficient to ensure that the observed effect after that point will be the sustained long-term improvement without any Novelty Effect influence.

We also developed a Driver Response Model that combines a delayed Driver Awareness effect with the Novelty Effect to explain the driver response observed in the “after” period. A freeway experiment was then used to validate the Driver Response Model and illustrate the application of the framework in evaluating the effectiveness of larger signage (i.e., a larger Yield sign) and symbolic pavement markings (i.e., a Yield Line). Our framework allows for meaningful comparisons of these different strategies to identify the best one in the context specified for the analysis (e.g., midday suburban), using the Baseline Compliance Prediction Model to control for inevitable differences among the implementations (e.g., changes in traffic patterns that occur at only one of the experimental locations) and the Driver Response Curve to account for the fact that each strategy will have its own novelty response curve and will take a different amount of time to reach stability. The primary benefits of our approach are that it provides a well-defined method for separating the long-term effect of a compliance strategy from transient improvements, and that it enables us to quantify the strength and duration of the Driver Awareness and Novelty Effects.

One distinguishing feature of this study is its use of multiple sites to evaluate different strategies individually, which also enables us to incorporate unchanged control ramps into the analysis as well. Such control sites are important for removing the effects of external variables

that affect driver compliance, as evidenced by the fluctuations in the baseline predictions shown in Figures 4 and 19. Furthermore, the high degree of similarity among our study sites allows us to make more meaningful performance comparisons than are possible with other studies, as discussed in the “Literature Review” section.

Summary of Study Findings

The real-world experiment that we applied our framework to gave insight into the appropriate distributions for modeling the persistence of Novelty Effect (i.e., a Normal distribution) and the Driver Awareness effect (i.e., a Normal distribution as well), as well as which factors are significant in predicting the compliance of drivers in a freeway environment (e.g., average freeway speed and prevalence of sunny weather).

Using the unified results from our framework, we found that a larger Yield sign at a freeway entrance ramp resulted in a 22% reduction in violations, with a 95% confidence interval of 10% to 26%. We also found that the Novelty Effect (if present) can last much longer than the 4-8 week duration that current literature assumes, with stability occurring after about 16 weeks. Furthermore, our data and our fitted Awareness Curve suggested that it took drivers an average of 1.6 weeks to notice and properly interpret the change, which is sensible given the straightforward nature of a larger sign. The overall Driver Response Curve for the larger sign, based on the unification of results from the Yield Line and Larger Sign data, is shown in Figure 28.

For symbolic pavement markings (i.e. a Yield Line), the Novelty Effect had a statistically significant influence on the Driver Response Curve, with an even longer persistence than in the case of the larger sign. Specifically, the Novelty Effect did not disappear until about 25 weeks after the Yield Line was installed, with a pronounced peak in its effect at about Week 10. Our fitted Awareness Curve suggested that it took drivers an average of 5.4 weeks to notice and properly interpret the strategy—over three times longer than it took for the larger Yield Sign, but reasonable given the limited deployment of Yield Lines in the field to date (making them less familiar to drivers, which could increase the time it takes for them to properly interpret its meaning). Overall, we found the Yield Line to be 27% effective according to our MOE (see Equation 6), with a 95% confidence interval of 17% to 34%. The overall Driver Response Curve for the Yield Line, based on the unification of results from the Yield Line and Larger Sign data, is shown in Figure 30.

We ended by exploring how our proposed framework compared to a conventional before-and-after approach. The conventional approach includes no method for excluding the initial, unstable portion of the driver response curve, and since there is currently no data to inform how long the instability lasts after implementing a strategy, agencies would either need to make unreliable guesses about how long they should wait, or simply choose to begin the “after” analysis immediately after the change is made—the prevalent approach in literature at the moment. We examined the differences in overall MOE estimates from the conventional method and our framework, and discussed the effects of failing to take into account the Novelty Peak and Awareness Effect. Depending on the particular assumptions made for the conventional approach, our MOE estimates varied greatly, ranging from 8.7% to 26.3% for the Larger Sign analysis and from -0.8% to 39.7% for the Yield Line analysis.

Comparisons to Other Regions

To gain insight into how much the effectiveness of a particular strategy may differ when the demographics of the user pool change, we also investigated the feasibility of implementing and analyzing a compliance strategy at a similar yield-controlled on-ramp in a different part of Los Angeles. For this, we considered either the Alameda St entrance to southbound Interstate 5 in Burbank (see Figure 36), or the similarly-designed Western Ave entrance about one mile farther south along the freeway.

Although these arterial interchanges use the same compressed cloverleaf design found along the Interstate 10 corridor, the two southbound entrances are the only ones with Yield signs, meaning that our control site options would be limited to one location. As we found in our analysis of the Larger Sign and the Yield Line along Interstate 10, some control locations are better predictors of behavior at the experimental location than others. Although we were able to test different control sites and select the ones that enabled us to make the best predictions for behavior at the experimental ramps for our research along Interstate 10, the lack of options for control sites in Burbank would limit our ability to develop accurate prediction models for the behavior of drivers on those ramps, which could ultimately lead to less accurate MOE estimates.

More important, however, is that the surrounding terrain at the Western Ave and Alameda St freeway entrance ramps is not conducive to observing driver behavior from adjacent private property as we have done along Interstate 10. The freeway is separated from the surrounding neighborhood by a berm on Caltrans' right-of-way (see Figure 36) with a sound wall on its crest, such that looking over it from adjacent private property would require camera stands much taller than the 15 feet of our current design (due to the ramp merge areas being immediately on the other side of the wall). Since it would not be feasible to obtain permits to observe the freeway from the Caltrans right-of-way (even from the neighborhood side of the sound wall), the lack of viewing locations on private property or nearby pedestrian-accessible overcrossings meant that there would be no way to collect video data at these locations. Consequently, we were unable to perform any analyses of strategies along Interstate 5 as we have done along Interstate 10.

Applications

A hallmark feature of this framework is that it enables us to compare different compliance strategies to identify the best one in a given context (e.g., midday suburban driving conditions). Not all sites with compliance issues are well suited for this framework since not all sites have good candidate control locations available, but those that do may be used to test several different strategies (i.e., one strategy at each site, leaving some control sites unchanged), so that the resultant improvements from each strategy can be compared and general conclusions drawn about the relative effectiveness of each. When a compliance issue arises at a unique location, the framework cannot be applied directly (i.e., to analyze the site itself) due to a lack of control locations, but the results of this framework's application to other sites that were more suitable (i.e., sites with comparable control sites available, as we had here) can still be used to inform the best strategy to implement at the unique location.

Because this framework compares the effectiveness of different strategies that have been implemented on a set of inherently similar sites, it can be used to make meaningful comparisons of strategies while avoiding the bias that can occur when comparing one strategy implemented at a low-compliance site with another strategy implemented at a high-compliance site. Another benefit to this framework is that any biases in compliance measurements (e.g., a tendency to

*Ramp layout and features at Alameda St
(Western Ave interchange uses a similar design)*

*Environment near the yield point at Alameda St,
on the neighborhood side of the sound wall.*

**FIGURE 36: Site characteristics of the compressed cloverleaf interchanges
along Interstate 5 in Burbank, CA.**

undercount certain instances of compliance) will occur at all sites, so that it does not compromise the conclusions regarding the relative performance of each strategy.

Another asset of this study is that the compliance strategies being tested are not unique to Yield signs, and can be applied to other situations where non-compliance is a problem. This will become even more relevant as agencies begin implementing advanced traffic management systems like HOT lanes and variable speed limits whose effectiveness depends heavily on driver awareness and compliance. However, these strategies also have clear relevance to traditional compliance problems like ignorant violators of “No Turn On Red” restrictions, unmarked pedestrian crosswalks, and static speed limits (especially at locations where the limits change).

Knowledge of the characteristics of the unstable acclimation period, including Driver Awareness and Novelty Persistence distributions, can enable researchers to predict drivers’ responses to other changes in their environment—for instance, how long it will take drivers to respond to roadside Changeable Message Signs at long-term construction sites. Such knowledge may also be insightful to other traffic situations that do not involve compliance, such as the level of utilization of a new parking structure on a college campus; in this case, it may take days or weeks for drivers to realize the structure has opened, at which point the usage of the lot may peak. As the new lot becomes oversubscribed and drivers begin to realize that it is no longer more convenient than their original lot (despite being new), the novelty wears off and usage levels of the new lot stabilize at a lower, long-term level. Having information about the basic shapes of Awareness and Novelty Persistence curves can help agencies predict when such a peak might occur at a new facility based on just a few initial data samples, as well as how long it might last.

As this framework is applied to analyze additional compliance strategies and in other contexts, one potential benefit to practitioners is that they will begin to have a growing pool of literature on the duration of the Novelty Effect and Driver Awareness effect for various compliance strategies, which would enable them to estimate how long it may take for conditions to stabilize as they implement similar strategies in their own jurisdictions. From this, for example, researchers may begin to establish general guidelines for the “waiting period” that must occur prior to collecting “after” data for different kinds of compliance strategies (or other changes). After following such a “waiting period,” practitioners could be more confident that the driver response has stabilized, such that they may be able to conduct a more simplified (and conventional) before-and-after comparison study to analyze the effect that the strategy or change has had. However, without data on Novelty Persistence and Driver Awareness, which our framework provides, such guidelines for “waiting periods” would not be possible.

Future Work

At different times throughout 2010, Caltrans began upgrading the lighting at some of the locations on our corridor. The existing conditions at these locations initially had little to no operational lighting at the merge point, such that it was dark at night except for the spots illuminated by vehicle headlights. These ramp rehabilitations resulted in brighter lighting at the merge points and consequently made the existing Yield signs easier to see at night. To illustrate the difference, Figure 37 shows the lighting conditions before and after the rehabilitation of Location 6. These improvements also helped drivers anticipate what lay ahead on the ramp, allowing them to shift their attention from the pavement in their headlights to their overall surroundings—giving drivers more time to notice the Yield sign and potentially leading to fewer cases where it went unseen.

Before ramp rehabilitation.

After ramp rehabilitation.

FIGURE 37: Improvement in ramp illumination following rehabilitation at Location 6.

Such a strategy could be used for other places where night violations are an issue, such as at a dark rural intersection where traffic frequently violates a stop sign because it is simply not noticed in time (or at all). Since the weekly data we collected along Interstate 10 spans a 24-hour period from Saturday evening to Sunday evening, capturing the effects of a night-only strategy (e.g., improved lighting) would not be a problem. We might expect the duration of the Awareness effect to be shorter than that for the Larger Sign, as lighting improvements are more obvious and may not take two or three passes through the site to notice, though an analysis of the data will confirm or refute this. We can also compare the long-term effectiveness of better lighting to the effectiveness of our other compliance strategies (i.e., a larger Yield sign, a Yield Line, and “Yield” pavement lettering once analyzed). However, since the effect of our compliance strategies can be expected to vary if the amount of ambient light changes significantly (such as the change from midday to night), we will want to evaluate MOEs for those strategies using night data instead of midday data prior to making comparisons to the effectiveness of improved lighting. The strategy that performs best in brighter environments (e.g., midday) may not be the best performer in darker conditions. For example, a flashing beacon may be the most effective compliance strategy at night, whereas pavement markings may perform just as well during the day.

Time of day may play yet another role in determining the effectiveness of a given strategy, due to changes in driver attentiveness throughout the day (e.g., between early morning and late afternoon). For example, a subtle compliance strategy (like a larger sign) may have a measurable effect on driver behavior during midday hours, but it may take something with a stronger presence (e.g., an LED-illuminated sign) to affect early-morning drivers who may not be as alert or attentive. By looking at data from early-morning periods of the day, we would be able to test this idea to see if certain strategies become relatively less effective earlier in the day.

Such Time-of-day information would be useful to agencies trying to identify the best compliance strategy based on the context of the problem they are addressing. For instance, a compliance problem that occurs largely at night would require a strategy that works well in darker environments, whereas a situation involving predominately midday collisions might be addressed just as effectively by a simpler, less expensive strategy. Expanding our research to examine each strategy’s long-term effectiveness (β'_a) for different times of the day (e.g., morning, midday, dusk, and night) would provide results that could be used by agencies to identify the best strategies for the specific conditions of the compliance problems they want to address.

In more general terms, the type of driver response we presented here (i.e., the synthesis of Driver Awareness and Novelty Effects to produce the response curves we saw in Figures 16, 21, 26, and 30) may have applications to other kinds of changes in the driving environment apart from compliance strategies, such as volumes on a newly-opened facility, though these potential applications will need to be evaluated through additional research. Furthermore, although we might expect the compliance strategies we have analyzed here to perform similarly in other situations (e.g., for turn prohibitions or at unmarked pedestrian crossings), additional research will be needed to confirm or refute this.

Availability of Raw Data

To advance the breadth and depth of knowledge in the field of transportation engineering and to encourage additional research into the topics we have explored here, we are making the video data collected as part of this dissertation available for other researchers to use through December 31, 2015. Refer to Figures 4-6 for the details of what data are available. To request any, contact the author by email at *freeway@berkeley.edu* or *freeway@ucla.edu*.

8. REFERENCES

1. Garber, Nicholas J., and Srivatsan Srinivasan. *Effectiveness of Changeable Message Signs in Controlling Vehicle Speeds in Work Zones, Phase II*. Virginia Transportation Research Council, Charlottesville, VA, December 1998.
2. Peters, Mary. *Performance and Accountability Report for Fiscal Year 2007*. US Department of Transportation, Washington DC. 2007, pp. 112-113.
3. Hanscom, Fred R. *Evaluation of the Prince William County Collision Countermeasure System*. Virginia Transportation Research Council, Charlottesville, VA, February 2001.
4. Wang, Chunyan., Karen Dixon, and David Jared. *Evaluating Speed-Reduction Strategies for Highway Work Zones*. Transportation Research Record 1824, Paper No. 03-4099.
5. De Waard, Dick, Maaïke Jessurun, and Frank J. J. M. Steyvers. *Effect of road layout and road environment on driving performance, drivers' physiology and road appreciation*. Ergonomics, Volume 38, No. 7. 1995, pp. 1395-1407.
6. Morgan, J. F., A. R. Duley, and P.A. Hancock. *Driver responses to differing urban work zone configurations*. Accident Analysis and Prevention, No. 42. 2010, pp. 978-985.
7. Yan, Xuedong, Essam Radwan, Dahai Guo, and Stephen Richards. *Impact of "Signal Ahead" pavement markings on driver behavior at signalized intersections*. Transportation Research Part F 12, 2009, pp. 50-67.
8. Wasserman, Larry. *All of Statistics: A Concise Course in Statistical Inference*. Springer Science and Business Media Inc., New York, NY. 2004, pp. 363-364.
9. Pindyck, Robert, and Daniel Rubinfeld. *Economic Models and Economic Forecasts*. Irwin McGraw-Hill, Boston, Massachusetts. 1998 (4th Edition), pp. 164-166.
10. Green, Mark. *"How Long Does It Take to Stop?" Methodological Analysis of Driver Perception-Brake Times*. Transportation Human Factors, 2(3). 2000, pp. 195-216.
11. Taoka, G.T. *An Analytical Model for Driver Response*. Transportation Research Record 1213, Washington, DC., 1989, pp. 1-5.
12. Inman, Vaughan, and Gregory Davis. *The Effects of In-Vehicle and Infrastructure-Based Collision Warnings at Signalized Intersections*. Federal Highway Administration Publication FHWA-HRT-09-049, December 2009.
13. Kahane, Charles J., and Ellen Hertz. *The long-term effectiveness of center high mounted stop lamps in passenger cars and light trucks*. National Highway Traffic Safety Administration, Washington D.C., March 1998.
14. *Manual on Uniform Traffic Control Devices*. FHWA, 2003 Edition 1 Revision, Section 2B.

15. *California Intersection Decision Support: A Systems Approach to Achieve Nationally Interoperable Solutions II*. Task Order 5600 Final Report. p. 68.

APPENDIX A:

Initial research proposal submitted to Caltrans in November 2008 describing the prevalence of the compliance problem and the scope of the project to address it.

University of California
Department of Civil and Environmental Engineering
TRAFFIC SAFETY STUDY PROPOSAL

TO:

Date: November 26, 2008

FROM:

Robert Campbell
Department of Civil and Environmental Engineering
University of California, Berkeley

SUBJECT:

Request for Authorization to Proceed With Study

BACKGROUND. The San Bernardino Freeway through the San Gabriel Valley (in eastern Los Angeles) adopted its right-of-way from the former Ramona Expressway, and in doing so, inherited a number of compressed cloverleaf interchanges between Interstate 710 and SR 19 (see Figure 1). The design of these interchanges encourages a number of risky driver behaviors including failure to yield at merge points (see Figure 2), leading to flow disruption and high risk of collisions. Indeed, as a Los Angeles resident having personal experience with these interchanges, I can validate the statement that they are some of the most dangerous ones out of the entire metropolitan area (the Pasadena Freeway aside).

Fortunately, there are three cost-effective, easy-to-implement solutions that can be investigated quickly and easily at any one of these interchange locations. Each solution could be installed, evaluated, and removed in a 24-hour time frame, with the installation and removal set for brief spans anytime between 1 and 3 a.m. to preclude any inconvenience to ramp users. The findings from the study could then be passed on to Caltrans and other transportation officials throughout the country to inform and guide their actions in addressing similar traffic safety problems at other locations.

Figure 1: Interstate 10 through the San Gabriel Valley, Los Angeles, CA

PROBLEM DESCRIPTION. From a preliminary corridor study conducted in March 2008, an analysis of hours of video data collected at various locations and times of day showed that 47% of merging vehicles failed to yield when appropriate. This number varied from 41% to 55% depending on the time and location, though this does not represent a statistically significant difference. It is therefore reasonable to conclude that the safety risks associated with this improper driver behavior are present at each street interchange throughout the corridor.

Though some of these offending drivers merely disrupted the flow of through traffic that had the right-of-way, others were at high risk of colliding with the through traffic due to the speed and timing of their failures to yield. In the video data collected, there was on average more than one near-collision per hour at these ramp merge points, where a near-collision required the driver with the right-of-way to brake hard and swerve to avoid impact.

Figure 2: Layout of interchanges and identification of high-risk ramp merge.

SOLUTIONS.

Larger Signs. At \$300, a larger sign would make for a very affordable and straightforward solution.

Background: Based on some simple surveying techniques, the size of the currently installed Yield Signs is estimated to be the minimum size allowed (for any use) by MUTCD standards, or 30 inches per side. MUTCD guidance for freeway installations is 60 inches per side, or double the dimensions of the currently installed signs along Interstate 10 (having four times as much display area as the current signs).

Implementation: A prototype 60-inch sign could be ordered and rigged with secure mounting clips on the back that allow it to attach securely (but non-permanently) to an existing 30-inch sign, making installation and removal as easy as tightening or loosening the clips.

Precedent: Large and even oversized regulatory signs are already in use along many of California's freeways and highways, such as the Stop Signs at many two-way stop intersections along US 395. Furthermore, in this study, the prototype sign used would be the size recommended by MUTCD, such that no standards are being violated.

Supplemental Pavement Marking. A simple Yield Line on the pavement offers more relevant placement than any posted sign, and has fewer visual distractions around it than a posted sign.

Background: Due to the geometry of the entrance ramps along Interstate 10, the Yield Signs were not installed immediately adjacent to the Yield Point. Instead, their placement and

orientation makes them difficult or impossible to see for the last 25–50 feet of the approach. Pavement markings emphasize the need for ramp traffic to yield at the merge point, and are more likely to be noticed by drivers since their eyes will naturally be on the roadway before them. MUTCD standards permit yield points of this type to be marked with both a broken line and accompanying “Yield” text.

Implementation: A Yield Line on the pavement would be simple to install on a temporary basis using white pavement marking tape. Black paint could be used to mask portions of the strip to simulate broken line, since one long strip would take less time to install and remove than many small cutouts. The tape is designed to be temporary and can be installed or removed quickly without the aid of special machinery.

Precedent: The approaches to the intersection between SR 1 and SR 19 in Long Beach, CA, all contain yield signs accompanied by the same type of pavement markings proposed here. MUTCD allows for yield points to be marked in this way. Pavement marking tape was used on SR 13 in Berkeley, CA, to create temporary striping in August 2008.

Supplemental Sign LEDs. Bordering the Yield sign with subtle flashing LEDs enhances its visibility and may reduce the likelihood that a driver will fail to yield because he did not see the sign.

Background: At high-risk locations, a recent FHWA-supported trend has been to supplement posted signs (such as Stop signs or Pedestrian Crosswalk signs) with subtle flashing LEDs to enhance visibility. FHWA mentions the benefits of LED retrofitting of signs in the Safety section of its 2007 Performance and Accountability Report. Such measures help draw attention to signs in an unobtrusive way. The approach is also intuitive: the design of the sign remains unchanged, meaning drivers are not exposed to unfamiliar and potentially-confusing information.

Implementation: A prototype can be assembled from a standard Yield Sign (such as the one used in the first part of the study) and a simple array of red LEDs set to flash according to the rhythm already used on stop signs with similar LED highlighting. This prototype could be battery-powered, and installed or removed as easily as the large yield sign described earlier. If this solution proves effective, solar panels can be used to power the sign independently.

Precedent: Stop signs in California along US 395 at Coso Junction already have solar powered LED highlighting.

UNIQUE OPPORTUNITY. The Interstate 10 corridor through the San Gabriel Valley represents a unique opportunity to evaluate a number of solutions one by one to see which enhances yielding behavior most effectively. Along this corridor, there are multiple sites of similar configuration and traffic volume but of different users, meaning that each solution can be implemented at a different location and the resultant improvements in safety compared among the various sites.

REQUEST. I request the opportunity to study the effectiveness of each of these potential solutions discussed earlier (as well as combinations of them) on the ramps at any three of the exits along Interstate 10 between Atlantic Blvd and San Gabriel Blvd (preferably Atlantic Blvd, Garfield Ave, and New Ave). Costs do not need to be covered, nor do I need to be compensated for my time. For each of the three exits, I request two 24-hour periods so that each side of the freeway may be studied. For about 15 minutes at the beginning and end of the 24 hour spans, I would need access to the ramp to install or remove the candidate solution being evaluated (for this reason, it may be ideal to have the 24 hour period begin at about 2 a.m.).

APPENDIX B:

Proposal for implementation of the second and third compliance strategies of the project,
submitted to Caltrans in December 2010.

TO: **Mr. Joseph Kibe**, Office of Traffic Investigation, Caltrans District 7
 FROM: **Robert Campbell**, PhD Candidate in Civil Engineering, UC Berkeley
 RE: **I-10 On-Ramp Traffic Safety Study**

Executive Summary

After lengthy data analysis, for the final part of the study the following two items are recommended:

1. For the SB Atlantic Blvd on-ramp to WB I-10, where the traffic safety problem is most severe, place a simple solar-powered flashing ball light above (or below) the Yield sign.
2. For the traffic safety problem at the NB New Ave on-ramp to EB I-10, paint a Yield line (or the word "Yield") at the Yield point.

This traffic safety project is supported by members of the Federal Highway Administration.

Justification

From my analysis, the failure-to-yield problem at Atlantic Blvd is more severe than at other ramps on I-10. **Hundreds of people** every day are not yielding, leading to the dangerous situations seen in the video clips (see Evidence section). Furthermore, in the video data I've analyzed, there are vehicle collisions on these ramps at a rate of about **1.8 collisions for every 24 hours** of video.

The situation at the New Ave interchange is similar, with hundreds of drivers each day failing to yield. However, the collision rate is not as high at this location.

Evidence

It's one thing to have the problem described to you, and another to see it occurring. I strongly urge you to take 60 seconds to view samples of the problem, which I have uploaded to www.robertjaycampbell.com/study (there are only two clips, and watching both takes less than a minute altogether). As you watch the videos, keep in mind that these are **not** extreme/rare cases—they are common events.

Standard Practice for District 7

The two recommended items are nothing new for District 7, which already uses both of these things on other on-ramps that have similar problematic merge points. For example, the 5th street on-ramp to I-110 in Downtown has a flashing light to warn drivers of a sudden, sharp ramp merge just like the ones on I-10. Furthermore, in 2008 Caltrans painted a Yield line at the Garfield Ave entrance to I-10 East, which is another interchange of similar design adjacent to the New Ave and Atlantic Blvd interchanges.

Lowering Cost

I realize that budgets are very tight these days, and I would be willing to contribute to the cost of both recommended strategies if it helps get the ball rolling more quickly on this (it helps you by making things safer on these ramps, and it helps me get closer to completing my PhD). I've saved up \$5000 for it, so if there's a way I could make a directed contribution to Caltrans or something for these two items, I would be glad to help cover the costs. As some very basic flashing LED traffic lights start at only \$70, I hope this helps makes these changes feasible.¹

¹ <http://www.trafficsafetystore.com/SubCatMProdDetails.aspx?CID=LED-Flasher-Lights&SCID=Solar-Barricade>

APPENDIX C:

Sample contract with a homeowner regarding the placement of equipment on his property to record the adjacent on-ramps for one year.

Agreement

The Resident (you)

I, _____ (the resident):

- Understand that this compensation **will be exactly \$500**, of which \$250 will be paid at the start of the data collection period and \$250 will be paid at the end.
- Allow the researcher (and grant permission for him) to collect freeway video data from *the southwest corner of my property (next to the small fruit tree)*. This also applies to his research assistants. I understand that he will notify me beforehand if on any particular day and time, anyone other than himself will be collecting data or working on the equipment.
- Will contact the researcher immediately if any equipment appears damaged, stolen, modified, or tampered with.
- Am not responsible for any theft or damage to the equipment being placed on my property.
- Understand that this data collection will take place **from June 1, 2011 to May 31, 2012**, and that all equipment will be removed by the end of this period.
- Understand that if for any reason on the part of the resident the equipment must be removed prematurely, I will be required to refund the original payment and will not receive the second (final) payment. I also understand that in this circumstance, I must also give the researcher advance notice of one month.
- Will make good faith efforts to accommodate the needs of the researcher (such as being reasonably available to talk to or meet with the researcher if it is needed).
- Will return the researcher's e-mails or phone calls (if any) within 1 week.
- Will allow the researcher to secure the equipment by *wrapping a cord around the base of the fruit tree*.

Signature

Date

The Researcher

I, Robert Campbell (the researcher):

- Agree to collect data of the freeway traffic only, and to be wholly respectful of the resident(s) and his (their) privacy.
- Assert that this research is being conducted with permission from and in cooperation with Caltrans and the University of California, Berkeley.
- Will set up the equipment in such a way that the data can be collected without disturbing the resident(s).
- Will notify the residents beforehand if anyone other than myself will be collecting data or working on the equipment.

Signature

Date

I-10 Ramp Study

Payment Confirmation and Contact Information

Payments Received

Please sign below to indicate receipt of payment:

Payment 1: \$250 (start of study)

Signature

Date

Payment 2: \$250 (end of study)

Signature

Date

Contact Information

For all questions, concerns, or problems, please do not hesitate to contact the researcher at:

Robert Campbell, PhD Student
UC Berkeley Institute of Transportation Studies
Department of Civil and Environmental Engineering
(323) 574-1208
freeway@ucla.edu

Additional Contacts

The following contacts are being provided for reference. Please **do not** contact them unless it becomes absolutely necessary.

Caltrans Contact

Joseph Kibe
Caltrans District 7
Office of Traffic Investigation
(213) 897-5742
joseph_kibe@dot.ca.gov

UC Berkeley Supervising Professor

Professor Alex Skabardonis
UC Berkeley Institute of Transportation Studies
Department of Civil and Environmental Engineering
(510) 642-9166
dromeas@berkeley.edu

APPENDIX D:

Instructions for installation and use of the Windows Video Processing Tool
to facilitate the reviewing of footage and the encoding of yield events.

Windows Video Processing Tool

QuickTime 7

The first thing you will need, if you don't have it already, is QuickTime 7. If you don't already have it on your computer, either in your Applications folder or your Utilities folder, you can download it from Apple here:

<http://www.apple.com/quicktime/download/>

Video Processing Tool

Next, you will need to download and install a program (under 2 MB) I've built that will let you control the video footage easily using the keyboard, and will also make it easy to record proper yields and failures to yield as they occur. The program installer can be downloaded here:

http://www.robertjaycampbell.com/Spring_2012/Windows/Tool.zip

Just unzip the folder, run Setup.exe, and launch the program if it doesn't automatically open itself.

A brief description of how the program works:

1. You can control the video using the options in the *Movie* menu.
 - a. To jump to a certain time in the video file, use the "Go to Time" option.
 - b. To speed up the video in increments of 0.25x, use the "Faster" option.
Note that there is a keyboard shortcut associated with this action as well (Alt + Right Arrow). The current playback speed is reported in the status bar at the bottom of the window. The "Slower" option behaves analogously.
 - c. To pause or unpause the video, use the button in the main viewer window. The keyboard shortcut for this is the spacebar.
 - d. To jump back 5 seconds in the video, use the "Back 5 Sec" option. The corresponding keyboard shortcut is Alt + Down Arrow. The "Forward 5 Sec" option behaves analogously.
2. To mark an event, just use the "Proper Yield" and "Failure to Yield" buttons in the main viewer window. There are *Alt key* shortcuts you can use for these buttons as well, as indicated by the underlined letter on each button when holding Alt. For example, you can mark a Proper Yield using Alt + Y. The most recent event you marked shows up in the status bar at the bottom of the window.
3. The program will create a file on your desktop called "output.txt" and will record the file name, timestamp, and event for all Proper Yields and Failures to Yield

- that you mark. If the file already exists, it will be amended. Try to keep all the Yield data for a particular week in the same output file, and then rename the file to the week it represents once you're finished (for example, put all the Yield Data for footage from "Oct 18 2009" in the same text file, and rename that file to "Oct 18 2009" when you're finished.
4. If you need to correct something in the output file, make a note to yourself in a separate text file and then make the corrections after closing the Yield application. Do not try to edit the file while the Yield application is running.

APPENDIX E:

Instructions for installation and use of the Mac OS X Video Processing Tools
to facilitate the reviewing of footage and the encoding of yield events.

Mac OS X Video Processing Tools

QuickTime 7

The first thing you will need, if you don't have it already, is QuickTime 7. If you don't already have it on your computer, either in your Applications folder or your Utilities folder, you can download it from Apple here:

<http://support.apple.com/kb/DL923>

Keyboard Shortcuts

Next, you will need to download and install a free program called Spark, which lets you control applications like QuickTime using keyboard shortcuts. You can download it here:

<http://www.macupdate.com/app/mac/14352/spark>

Downloading QuickTime Scripts

You'll also need to download some scripts that will let you control QuickTime. The scripts can be downloaded from here:

http://www.robertjaycampbell.com/Spring_2012/Mac/Scripts.zip

A brief description of the scripts you'll find:

1. **Play-Pause** — Will pause, or resume playback at normal speed.
2. **Back 5 Seconds** — Will take you back 5 seconds and begin playback at normal speed. You can change the amount of time it skips back by opening the script file and changing the variable at the top.
3. **Forward 5 Seconds** — Same idea as the "Back 5 Seconds" script.
4. **Slow Down** — By default, this script slows down the video in increments of 0.5x (so running it twice will cause the video to completely stop assuming you started at normal speed, and running it even more will cause the video to play in reverse). You can change the increment amount by opening the script file and changing the variable at the top.
5. **Speed Up** — Same idea as the "Slow Down" script.

Configuring the shortcuts

Once you've installed Spark, you'll need to configure it. We want to set it up so that the scripts above can be invoked using keyboard shortcuts. In general, I recommend the following shortcuts:

1. **Command + Control + Option + Left Arrow** for the "Slow Down" action.
2. **Command + Control + Option + Right Arrow** for the "Speed Up" action.

3. **Command + Control + Option + Up Arrow** for the “Forward 5 Seconds” action.
4. **Command + Control + Option + Down Arrow** for the “Back 5 Seconds” action.
5. **Command + Control + Option + Space Bar** for the “Play/Pause” action.

To configure Spark:

1. Launch Spark and choose **File ► New HotKey ► AppleScript**.
2. In the panel that opens up, you'll see a section near the top that says **Shortcut**. Click the area to the right that says "click to edit". It should now say "Type hotkey", at which point you can enter the keyboard shortcut you want to use for the new action—For example, **Command + Control + Option + Left Arrow** for the "Slow Down" action.
3. Give the new action some sort of descriptive name in the **Name** text box, like “Slow Down”.
4. Toward the middle of the panel, you'll see two tabs: **Source** and **File**. Click on the **File** tab, press the **Choose** button, and select the script file associated with the action you want to perform (for example, the “Slow Down” script file).
5. Finish by clicking **Create** in the lower right corner.
6. Repeat for all other actions (e.g., Speed Up, Back 5 Seconds, etc.)
7. Click the **Start Spark Daemon** button in the main Spark window to make the keyboard shortcuts active.
8. Quit Spark (don't worry, the keyboard shortcuts will remain active).

When you're finished working and want to disable the keyboard shortcuts, reopen Spark and click the button at the bottom of the window that says **Stop Spark Daemon**.

Installing the Yield App

Now that you can use keyboard shortcuts to remote-control QuickTime, the last thing to do is download the application that records Yields and Failures to Yield in the video file as you watch it. You can download the application here:

http://www.robertjaycampbell.com/Spring_2012/Mac/Yield_App.zip

Here are some important notes on using the Yield app:

1. The app will automatically put a file on your desktop called “output.txt” with a list of all the events you recorded. If the file exists on the desktop already, the output.txt file will simply be amended when you launch the Yield app. Try to keep all the Yield data for a particular week in the same output file, and then rename the file to the week it represents once you're finished (for example, put all

- the Yield Data for footage from “Oct 18 2009” in the same text file, and call that file “Oct 18 2009” when you’re finished.
2. If you need to correct something in the output file, make a note to yourself in a separate text file and then make the corrections after closing the Yield application. Do not try to edit the file while the Yield application is running.
 3. When you finish a video file, close the Yield application even if you plan on processing another video file. When you have the new video file open, relaunch the Yield application.
 4. Don’t run the Yield application without a video open in QuickTime. Also, don’t run it with more than one video open in QuickTime at once.

APPENDIX F:

Announcement seeking enthusiastic research assistants
to contribute to the video processing effort.

UC Berkeley
Institute of
Transportation Studies

Paid Positions for Undergraduate Research Assistant

Background

We are investigating the effectiveness of different compliance strategies. For example, when drivers don't yield to pedestrians in unmarked crosswalks, some common strategies include putting up Ped Xing signs, painting lines on the pavement, or adding a flashing beacon above the sign. In this research, we're looking at a compliance situation involving Yield signs, but the strategies are similar.

What You'd Be Doing

You would be reviewing freeway video footage and noting whether vehicles yield properly or not as they arrive at the Yield signs on different on-ramps in Los Angeles. Here are two example 30-second clips showing a very clear proper yield (the first clip) and a few equally clear improper yields (the second clip): <http://www.robertjaycampbell.com/study>

Hours and Compensation

Everything you would need to do can be done from home using your own computer, so you'd have the freedom to choose your own work hours. We will provide you with tools (Windows and Mac compatible) to make reviewing the video footage easier for you. The time commitment would be about 6 hrs/wk, though this is also pretty flexible—the critical factor is that you would need to have put in a minimum of approximately 40 hours by March 24, 2012. **For your work, you would be compensated \$15/hr.**

For More Information or to Apply

Interested students should email Robert Campbell at freeway@berkeley.edu with any questions or to apply.

There is no formal deadline, although prospective applicants should contact Robert as soon as possible as the positions will be filled on a rolling basis. When applying, please include a contact phone number, your year/department, and a copy of your schedule to help us find a time to meet with you on campus for an informal interview.

APPENDIX G:

Instructions for installation and use of the Windows Traffic Counts Tool
used to record on-ramp traffic volumes manually.

Windows Traffic Counting Tool

QuickTime 7

The first thing you will need, if you don't have it already, is QuickTime 7. If you don't already have it on your computer, either in your Applications folder or your Utilities folder, you can download it from Apple here:

<http://www.apple.com/quicktime/download/>

Video Processing Tool

Next, you will need to download and install a program (under 2 MB) I've built that will let you control the video footage easily using the keyboard, and will also make it easy to record red-path and yellow-path vehicles. The program installer can be downloaded here:

http://www.robertjaycampbell.com/Spring_2012/Counts_Program/Tool.zip

Just unzip the folder, run Setup.exe, and launch the program if it doesn't automatically open itself. If you previously installed the Video Processing Tool (for recording Yielding behavior), you will need to uninstall this one first using the Add/Remove Programs control panel.

A brief description of how the program works:

1. You can control the video using the options in the *Movie* menu.
 - a. To jump to a certain time in the video file, use the "Go to Time" option.
 - b. To speed up the video in increments of 0.25x, use the "Faster" option. Note that there is a keyboard shortcut associated with this action as well (Alt + Right Arrow). The current playback speed is reported in the status bar at the bottom of the window. The "Slower" option behaves analogously.
 - c. To pause or unpause the video, use the button in the main viewer window. The keyboard shortcut for this is the spacebar.
 - d. To jump back 5 seconds in the video, use the "Back 5 Sec" option. The corresponding keyboard shortcut is Alt + Down Arrow. The "Forward 5 Sec" option behaves analogously.
2. To mark an event, just use the "Red-Path Vehicle" and "Yellow-Path Vehicle" buttons in the main viewer window. There are *Alt key* shortcuts you can use for these buttons as well, as indicated by the underlined letter on each button when holding Alt. For example, you can mark a Red-Path Vehicle using Alt + R. The most recent event you marked shows up in the status bar at the bottom of the window.

3. The program will create a file on your desktop called “output.txt” and will record the file name, timestamp, and event for all Red-Path and Yellow-Path vehicles that you mark. If the file already exists, it will be amended. Try to keep all the Count data for a particular week in the same output file, and then rename the file to the week it represents once you’re finished (for example, put all the Count Data for footage from “Oct 18 2009” in the same text file, and rename that file to “Oct 18 2009” when you’re finished).
4. If you need to correct something in the output file, make a note to yourself in a separate text file and then make the corrections after closing the application. Do not try to edit the file while the application is running.

APPENDIX H:

Data tables for the analyses of the Larger Sign, the Ramp Rehabilitation, and the Yield Line.
Also includes hourly data for the “before” period at Location 8.

Compliance Rates and Covariate Data by week for Location 8 (Larger Sign)

Date	Compliance at Location 8*		Compliance at Location 5 Control Ramp		Compliance at Location 1 Control Ramp		Compliance at Ramp at Location 9		On-ramp volume in three hours		Natural Log of On-ramp volume		Freeway speed passing the on-ramp		Percent of time the freeway is slower than 55 mph		Percent of time the ramp flow exceeds 60 veh in 5 min		Percent of time the on-ramp flow exceeds 70 veh in 5 min		Percent of time the on-ramp flow exceeds 80 veh in 5 min		Percent of time the weather is sunny at Location 5	
	Location 8*	Compliance at Location 8*	Location 5 Control Ramp	Compliance at Location 5 Control Ramp	Location 1 Control Ramp	Compliance at Location 1 Control Ramp	Ramp at Location 9	Compliance at Ramp at Location 9	On-ramp volume in three hours	Natural Log of On-ramp volume	Freeway speed passing the on-ramp	Percent of time the freeway is slower than 55 mph	Percent of time the ramp flow exceeds 60 veh in 5 min	Percent of time the on-ramp flow exceeds 70 veh in 5 min	Percent of time the on-ramp flow exceeds 80 veh in 5 min	Percent of time the weather is sunny at Location 5								
Oct 18, 2009	45.8% (179)	51.2%	43.3%	52.3%	1876	7.537	67.298	0%	31%	17%	3%	94%												
Oct 25, 2009	47.0% (166)	57.6%	44.9%	54.1%	2003	7.602	67.990	0%	36%	19%	8%	100%												
Nov 1, 2009	47.0% (134)	56.3%	37.1%	49.6%	1835	7.515	68.561	0%	33%	14%	11%	100%												
Nov 8, 2009	49.4% (162)	50.7%	55.1%	53.8%	1900	7.550	64.636	11%	31%	8%	6%	100%												
Nov 15, 2009	47.6% (187)	51.1%	43.5%	48.5%	1816	7.504	67.910	0%	19%	3%	0%	100%												
Nov 22, 2009	46.9% (175)	51.7%	46.2%	50.8%	1726	7.454	68.330	0%	22%	14%	0%	100%												
Nov 29, 2009	41.7% (180)	53.9%	37.4%	51.4%	1771	7.479	69.258	0%	14%	3%	0%	100%												
Dec 6, 2009	49.4% (158)	51.6%	44.0%	48.6%	1798	7.494	68.114	0%	11%	0%	0%	42%												
Dec 13, 2009	43.7% (174)	44.9%	48.4%	46.5%	1758	7.472	73.691	0%	6%	0%	0%	97%												
Dec 20, 2009	49.4% (168)	41.8%	48.4%	60.1%	1813	7.503	69.177	6%	14%	3%	0%	78%												
Dec 27, 2009	44.4% (153)	55.3%	45.7%				73.633	0%				61%												
Jan 3, 2010	48.6% (177)	50.9%	49.2%				72.377	0%				81%												
Jan 10, 2010	53.0% (168)	52.3%	39.6%				72.768	0%				56%												
Jan 17, 2010	56.3% (158)	58.9%	50.9%				71.191	0%				44%												
Jan 24, 2010	50.0% (158)	53.2%	43.0%				71.942	0%				100%												
Jan 31, 2010	53.6% (166)	54.4%	42.5%				67.709	0%				100%												
Feb 14, 2010	60.8% (148)	57.9%	39.7%				71.267	0%				100%												
Mar 7, 2010	45.8% (190)	48.9%	39.0%				69.842	6%				69%												
Mar 14, 2010	52.4% (286)	59.0%	47.5%				71.065	0%				100%												
April 4, 2010	54.8% (177)	53.8%					70.425	0%				72%												
May 2, 2010	50.0% (184)	55.3%					36.731	89%				92%												
July 18, 2010	53.6% (235)	54.0%					73.195	0%				100%												

*Total number of observed yielding events (i.e., proper yields and failures to yield) shown in parentheses.

Compliance Rates and Covariate Data by week for Location 5 (Ramp Rehabilitation)

Date	Compliance at Location 5*	Compliance at Location 6 Control Ramp	Compliance at Location 1 Control Ramp	Compliance at Location 8 Control Ramp	On-ramp volume in three hours	Natural Log of On-ramp volume	Freeway speed passing the on-ramp	Percent of time the freeway is slower than 55 mph	Percent of time the ramp flow exceeds 60 veh in 5 min	Percent of time the ramp flow exceeds 70 veh in 5 min	Percent of time the on-ramp flow exceeds 80 veh in 5 min	Percent of time the weather is sunny at Location 5
Oct 18, 2009	51.2% (99)	73.2%	43.3%	45.8%	914	6.818	60.497	11%	6%	6%	3%	97%
Oct 25, 2009	57.6% (91)	71.2%	44.9%	47.0%	755	6.627	64.343	0%	0%	0%	0%	100%
Nov 1, 2009	56.3% (104)	68.1%	37.1%	47.0%	819	6.708	64.850	0%	3%	3%	3%	100%
Nov 8, 2009	50.7% (119)	75.7%	55.1%	49.4%	927	6.832	63.893	0%	3%	0%	0%	100%
Nov 22, 2009	51.7% (98)	72.7%	46.2%	46.9%	1008	6.916	64.151	0%	6%	0%	0%	100%
Nov 29, 2009	53.9% (119)	75.9%	37.4%	41.7%	824	6.714	65.136	0%	3%	3%	3%	100%
Dec 6, 2009	51.6% (111)	67.8%	44.0%	49.4%	968	6.875	63.745	0%	8%	6%	3%	42%
Dec 13, 2009	44.9% (112)	76.0%	48.4%	43.7%	1209	7.098	62.824	3%	0%	0%	0%	92%
Dec 20, 2009	41.8% (72)	77.4%	48.4%	49.4%	1138	7.037	61.994	8%	0%	0%	0%	83%
Aug 22, 2010	44.9% (72)	75.5%	43.0%		964	6.871	64.678	0%				100%
Aug 29, 2010	50.9% (75)	75.9%	36.3%		1051	6.957	59.712	11%				83%
Sept 5, 2010	47.7% (85)	77.3%	39.5%		1128	7.028	62.481	0%				100%
Sept 12, 2010	40.5% (79)	73.0%	46.6%		1101	7.004	59.371	14%				100%
Sept 19, 2010	40.0% (69)	68.4%	45.7%		1036	6.943	62.635	0%				100%
Sept 26, 2010	40.5% (271)	69.3%	45.0%		1959	7.580						100%
Oct 3, 2010	65.6% (111)	N/A	39.6%		1519	7.326						
Oct 10, 2010	45.0% (136)	N/A	45.8%		1439	7.272						
Oct 24, 2010	45.6% (75)	N/A	50.0%		1080	6.985	64.520	0%				
Nov 7, 2010	59.6% (64)	N/A	41.5%		1075	6.980	67.512	0%				
Nov 14, 2010	49.3% (129)	88.1%			1274	7.150	66.982	0%				100%
Nov 21, 2010	53.2% (109)	77.0%	49.5%		1215	7.102	66.220	0%				94%
Nov 28, 2010	53.8% (100)	74.1%			1137	7.036	64.578	0%				100%
Dec 5, 2010	49.6% (79)	85.2%	N/A		1087	6.991	63.737	0%				0%
Dec 12, 2010	39.6% (92)	79.0%	N/A		1083	6.987	36.386	28%				100%

*Total number of observed yielding events (i.e., proper yields and failures to yield) shown in parentheses.

Compliance Rates and Covariate Data by week for Location 5 (Yield Line)

Date	Compliance at Location 5*	Compliance at Location 6		Compliance at Location 1		Compliance at Location 8		On-ramp volume in three hours	Natural Log of On-ramp volume	Freeway speed passing the on-ramp	Percent of time the freeway is slower than 55 mph	Percent of time the on-ramp flow exceeds 60 veh in 5 min	Percent of time the on-ramp flow exceeds 70 veh in 5 min	Percent of time the on-ramp flow exceeds 80 veh in 5 min	Percent of time the weather is sunny at Location 5
		Control	Ramp	Control	Ramp	Control	Ramp								
Oct 18, 2009	51.2% (99)	73.2%				45.8%		914	6.818	60.497	11%	6%	6%	3%	97%
Oct 25, 2009	57.6% (91)	71.2%				47.0%		755	6.627	64.343	0%	0%	0%	0%	100%
Nov 1, 2009	56.3% (104)	68.1%				47.0%		819	6.708	64.850	0%	3%	3%	3%	100%
Nov 8, 2009	50.7% (119)	75.7%				49.4%		927	6.832	63.893	0%	3%	0%	0%	100%
Nov 22, 2009	51.7% (98)	72.7%				46.9%		1008	6.916	64.151	0%	6%	0%	0%	100%
Nov 29, 2009	53.9% (119)	75.9%				41.7%		824	6.714	65.136	0%	3%	3%	3%	100%
Dec 6, 2009	51.6% (111)	67.8%				49.4%		968	6.875	63.745	0%	8%	6%	3%	42%
Dec 13, 2009	44.9% (112)	76.0%				43.7%		1209	7.098	62.824	3%	0%	0%	0%	92%
Dec 20, 2009	41.8% (72)	77.4%				49.4%		1138	7.037	61.994	8%	0%	0%	0%	83%
Jan 9, 2011	41.8% (92)	85.3%				N/A		1187	7.079	61.318	8%				72%
Jan 16, 2011	40.8% (110)	81.4%				N/A		1220	7.107	62.516	3%				100%
Jan 23, 2011	43.6% (88)	90.1%				N/A		1194	7.085	63.979	0%				100%
Feb 6, 2011	44.0% (121)	71.4%				49.3%		1336	7.197	64.428	0%				100%
Feb 13, 2011	44.6% (96)	71.8%				38.6%		1366	7.220	63.326	0%				100%
Feb 20, 2011	53.9% (78)	98.6%				40.1%		1255	7.135	64.931	0%				94%
Feb 27, 2011	60.1% (62)	84.1%				41.4%		1216	7.103	61.921	0%				100%
March 6, 2011	57.3% (167)	73.5%				34.5%		1780	7.484						
March 13, 2011	56.3% (80)	68.9%				42.6%		1342	7.202						
March 27, 2011	53.7% (85)	80.6%				36.2%		1182	7.075						
April 10, 2011	69.5% (71)	83.2%				38.8%		1160	7.056						
April 24, 2011	69.1% (74)	75.3%				34.2%		1124	7.025						
May 8, 2011	49.4% (95)	86.9%				36.4%		1151	7.048						
May 22, 2011	43.3% (100)	80.2%				39.2%		1118	7.019						
June 5, 2011	46.2% (94)	83.3%				35.5%		1113	7.015						
June 26, 2011	51.8% (113)	81.4%				32.5%		1197	7.088						
Aug 21, 2011	47.7% (107)	88.5%				39.4%		1125	7.026						
Nov 27, 2011	59.4% (96)	81.6%				38.3%		1182	7.075						
Jan 8, 2012	50.4% (113)	77.2%				36.6%		1334	7.196						

*Total number of observed yielding events (i.e., proper yields and failures to yield) shown in parentheses.

Compliance Rates and Covariate Data by hour for Location 8 (Larger Sign)

Date	Hour	Compliance at Location 8*	Compliance at Location 5 Control Ramp	Compliance at Location 1 Control Ramp	Compliance at Location 9 Control Ramp	On-ramp volume in three hours	Natural Log of On-ramp volume	Freeway speed passing the on-ramp	Percent of time the freeway is slower than 55 mph	Percent of time the ramp flow exceeds 60 veh in 5 min	Percent of time the on-ramp flow exceeds 70 veh in 5 min	Percent of time the on-ramp flow exceeds 80 veh in 5 min	Percent of time the weather is sunny at Location 8
Oct 18, 2009	11 AM – Noon	39.7% (58)	64.3% (58)	45.3% (58)	59.5% (58)	541	6.293	68.854	0%	25%	17%	0%	100%
Oct 18, 2009	Noon – 1 PM	57.4% (47)	60.0% (47)	47.6% (47)	49.1% (47)	620	6.430	67.118	0%	25%	17%	8%	83%
Oct 18, 2009	1–2 PM	44.0% (75)	61.5% (75)	38.8% (75)	50.8% (75)	715	6.572	65.984	0%	42%	17%	0%	100%
Oct 25, 2009	11 AM – Noon	47.9% (48)	60.9% (48)	51.9% (48)	54.1% (48)	561	6.330	69.061	0%	17%	8%	0%	100%
Oct 25, 2009	Noon – 1 PM	37.3% (51)	66.7% (51)	42.0% (51)	50.7% (51)	587	6.375	68.412	0%	33%	8%	0%	100%
Oct 25, 2009	1–2 PM	53.7% (67)	61.3% (67)	43.4% (67)	61.3% (67)	855	6.751	66.549	0%	58%	42%	25%	100%
Nov 1, 2009	11 AM – Noon	48.4% (64)	63.6% (64)	31.6% (64)	54.4% (64)	623	6.435	68.444	0%	25%	17%	17%	100%
Nov 1, 2009	1–2 PM	54.7% (53)	60.0% (53)	36.9% (53)	50.0% (53)	636	6.455	68.337	0%	33%	8%	0%	100%
Nov 8, 2009	11 AM – Noon	46.3% (41)	62.5% (41)	46.7% (41)	57.7% (41)	628	6.443	68.624	0%	33%	8%	8%	100%
Nov 8, 2009	Noon – 1 PM	56.5% (62)	57.7% (62)	50.3% (62)	51.7% (62)	643	6.466	61.083	25%	25%	8%	8%	100%
Nov 8, 2009	1–2 PM	41.7% (60)	64.5% (60)	40.2% (60)	52.0% (60)	629	6.444	64.639	8%	33%	8%	0%	100%
Nov 15, 2009	11 AM – Noon	68.4% (38)	62.5% (38)	51.3% (38)	57.4% (38)	579	6.361	68.321	0%	8%	0%	0%	100%
Nov 15, 2009	Noon – 1 PM	38.6% (70)	62.5% (70)	36.5% (70)	49.0% (70)	599	6.395	67.311	0%	25%	0%	0%	100%
Nov 15, 2009	1–2 PM	45.6% (79)	62.5% (79)	43.7% (79)	35.3% (79)	638	6.458	68.105	0%	25%	8%	0%	100%
Nov 22, 2009	11 AM – Noon	51.1% (45)	59.4% (45)	37.3% (45)	47.6% (45)	561	6.330	68.301	0%	25%	17%	0%	100%
Nov 22, 2009	Noon – 1 PM	43.4% (53)	39.5% (53)	46.3% (53)	56.5% (53)	538	6.288	68.683	0%	0%	0%	0%	100%
Nov 22, 2009	1–2 PM	46.8% (77)	57.8% (77)	49.4% (77)	47.6% (77)	627	6.441	68.007	0%	42%	25%	0%	100%
Nov 29, 2009	11 AM – Noon	42.9% (56)	69.6% (56)	37.0% (56)	54.8% (56)	531	6.275	69.681	0%	17%	0%	0%	100%
Nov 29, 2009	Noon – 1 PM	39.1% (64)	64.0% (64)	40.1% (64)	58.3% (64)	613	6.418	68.809	0%	17%	0%	0%	100%
Nov 29, 2009	1–2 PM	45.0% (60)	73.1% (60)	35.7% (60)	41.7% (60)	627	6.441	69.290	0%	8%	8%	0%	100%
Dec 6, 2009	11 AM – Noon	57.7% (52)	69.6% (52)	39.7% (52)	42.1% (52)	542	6.295	67.897	0%	0%	0%	0%	8%
Dec 6, 2009	Noon – 1 PM	36.4% (44)	81.8% (44)	53.4% (44)	64.1% (44)	600	6.397	68.017	0%	8%	0%	0%	50%
Dec 6, 2009	1–2 PM	51.6% (62)	77.4% (62)	40.9% (62)	43.1% (62)	656	6.486	68.430	0%	25%	0%	0%	67%
Dec 13, 2009	11 AM – Noon	37.5% (72)	64.7% (72)	43.6% (72)	45.7% (72)	588	6.377	75.288	0%	8%	0%	0%	92%
Dec 13, 2009	Noon – 1 PM	41.7% (48)	71.0% (48)	44.3% (48)	45.8% (48)	556	6.321	71.901	0%	0%	0%	0%	100%
Dec 13, 2009	1–2 PM	54.5% (55)	33.3% (55)	40.2% (55)	47.8% (55)	614	6.420	73.963	0%	8%	0%	0%	100%
Dec 20, 2009	11 AM – Noon	53.4% (58)	81.3% (58)	52.6% (58)	60.9% (58)	577	6.358	71.326	0%	0%	0%	0%	83%
Dec 20, 2009	Noon – 1 PM	44.4% (45)	45.5% (45)	45.7% (45)	59.1% (45)	573	6.351	74.278	0%	17%	0%	0%	75%
Dec 20, 2009	1–2 PM	50.0% (66)	54.2% (66)	47.9% (66)	60.3% (66)	663	6.497	62.956	17%	25%	8%	0%	75%

*Total number of observed yielding events (i.e., proper yields and failures to yield) shown in parentheses.

APPENDIX I:

Details of the procedure used to identify the best Choice Models for Locations 5 and 8, including intermediate results from each step of the process.

Choice Model Approach for Larger Sign at San Gabriel Blvd

Choice Models, Set 1:

For our first round of choice models, we allowed the following variables into the model, meaning that all of our models from Set 1 have six covariates and a constant term:

1. One of any of the freeway speed variables.
2. One of any of the on-ramp flow variables.
3. The weather variable.
4. The same-interchange control ramp yield rate.
5. Both of the same-side control ramp variables.

We tested all possible combinations of these variables to identify the one that fit our data the best, as indicated by the highest Rho-Squared value. Rho-Squared is given by:

$$\rho^2 = 1 - \frac{LL(Model)}{LL(Null Model)}$$

Since the denominator is the same for all of the models tested in this set, finding the model with the highest Log Likelihood (or the lowest -2LL, where LL represents the Log Likelihood) is equivalent to finding the model with the highest Rho Squared.

As the highlighting in Table 1 shows, the best model in this respect was #3. We will next look at the coefficients for Model #3 to identify which, if any, of the variables may not be significantly contributing to the performance of the model.

Table 1: Performance of Choice Models, Set 1

Model ID (for reference)	Speed Variable Used	On-ramp Flow Variable Used	-2 Log Likelihood
1	Freeway Speed	On-Ramp Flow	2317.848
2	Freeway Speed	Natural Log of On-Ramp Flow	2317.786
3	Freeway Speed	On-Ramp Flow above 60 v/5m	2317.125
4	Freeway Speed	On-Ramp Flow above 70 v/5m	2318.421
5	Freeway Speed	On-Ramp Flow above 80 v/5m	2317.526
6	Fwy Speed below 55	On-Ramp Flow	2319.939
7	Fwy Speed below 55	Natural Log of On-Ramp Flow	2319.939
8	Fwy Speed below 55	On-Ramp Flow above 60 v/5m	2318.995
9	Fwy Speed below 55	On-Ramp Flow above 70 v/5m	2320.666
10	Fwy Speed below 55	On-Ramp Flow above 80 v/5m	2319.527

Note that Table 1 does not indicate all of the predictor variables used in each model—it simply specifies which ones were used for those categories (e.g., freeway speed) where there were multiple predictor variables to choose from.

Best Candidate Choice Model so far

After evaluating 10 different models, we found the best one in terms of maximum likelihood to be Model #3. The details of this model are summarized here.

Choice Probabilities

The probability of a particular driver failing to yield is given by:

$$P(\text{Yielding} = 0) = \frac{e^0}{e^0 + e^{V(1)}}$$

$$\begin{aligned} V(1) = & 0.874 - 0.015(\text{Fwy Speed}) + 0.128(\text{OnRamp Flow} > 60) \\ & - 0.049(\text{Sunny weather}) \\ & - 0.729(\text{Same Interchange Control Hourly Yield Rate}) \\ & - 0.521(\text{Same Side Control Hourly Rate at New Ave}) \\ & + 1.436(\text{Same Side Control Hourly Rate at Atlantic Blvd}) \end{aligned}$$

The probability of a driver yielding properly is just $P(\text{Yielding} = 1) = 1 - P(0)$.

Table 2: Coefficient Details for Model #3

The coefficients have the following properties:

Variable	Coefficient	Standard Error	Significance
On-Ramp Flow is above 60 veh/5 min	.128	.113	.254
Freeway Speed passing on-ramp	-.015	.011	.163
Weather is sunny	-.049	.178	.782
Hourly yield rate at Same-Interchange Control Ramp	-.729	.751	.332
Hourly yield rate at Same-Side Control Ramp at New Ave	-.521	.505	.303
Hourly yield rate at Same-Side Control Ramp at Atlantic Bl	1.742	.977	.075
Constant	.874	1.015	.389

Based on the level of significance for these coefficients, we will next test six different versions of this model, each having one of the above insignificant (at a 10% level) variables removed. We recognize that doing so will cause a drop in performance with respect to Log Likelihood, but removing insignificant variables may also improve (i.e., decrease) the risk associated with the model, which is why we test these modified versions of Model #3 with variables removed.

Choice Models with Selected Variables Excluded

In an attempt to find a choice model that has a lower risk than that of our original Model #3, we develop the following six modified models that each have one of the six potentially insignificant variables removed (thus, each model in Table 3 has six parameters remaining in it). In each case,

the models have been optimized (i.e., the parameters that maximize the likelihood function have been used) for the subset of variables present in the model.

Table 3: Performance of Choice Models with One Variable Removed:

Model ID	Fwy Speed Variable Used?	On-Ramp Flow Variable Used?	Weather Variable Used?	Same-Interchange Control Ramp Used?	New Ave Control Ramp Used?	Atlantic Blvd Control Ramp Used?	Constant Used?	-2 Log Likelihood
13		✓	✓	✓	✓	✓	✓	2319.084
14	✓		✓	✓	✓	✓	✓	2318.426
15	✓	✓		✓	✓	✓	✓	2317.202
16	✓	✓	✓		✓	✓	✓	2318.066
17	✓	✓	✓	✓		✓	✓	2318.186
18	✓	✓	✓	✓	✓	✓		2317.869

As the highlighting in the above table indicates, the best model identified when one low-significance variable was removed was Model #15. Note that the Atlantic Blvd control ramp was included in every model of Table 3, since it was found to be significant at a 10% level in our initially selected optimal model (i.e., Model #3).

Next, we look at versions of our original model with two variables removed at a time instead of just one. Again, we recognize that this will negatively affect the log likelihood, but that it also has the potential to improve the risk associated with the model. Since this is more important to us, as better predictions are of greater priority than goodness of fit, we proceed now with an analysis of these new models that have two low-significance variables removed. All possible combinations of removed variables are tested, and the model with the best likelihood result is selected as our optimal model out of all those with two variables removed (i.e., five variables included).

Table 4: Performance of Choice Models with Two Variables Removed:

Model ID	Fwy Speed Variable Used?	On-Ramp Flow Variable Used?	Weather Variable Used?	Same-Interchange Control Ramp Used?	New Ave Control Ramp Used?	Atlantic Blvd Control Ramp Used?	Constant Used?	-2 Log Likelihood
21			✓	✓	✓	✓	✓	2320.805
22		✓		✓	✓	✓	✓	2319.119
23		✓	✓		✓	✓	✓	2319.902
24		✓	✓	✓		✓	✓	2319.908
25		✓	✓	✓	✓	✓		2319.248
26	✓			✓	✓	✓	✓	2318.452
27	✓		✓		✓	✓	✓	2319.242
28	✓		✓	✓		✓	✓	2319.502
29	✓		✓	✓	✓	✓		2319.399
30	✓	✓			✓	✓	✓	2318.159

31	✓	✓		✓		✓	✓	2318.19
32	✓	✓		✓	✓	✓		2317.869
33	✓	✓	✓			✓	✓	2318.996
34	✓	✓	✓		✓	✓		2318.456
35	✓	✓	✓	✓		✓		2318.389

In the case of two low-significance variables removed, the best model we found was #32 above. Recall that the Atlantic Blvd control ramp is included in every model of Table 4, since it was significant at a 10% level in our initially selected optimal model (i.e., Model #3).

We now consider three additional sets of choice models, all based on our original optimal model #3. In these additional sets, we have tried removing even more combinations of variables that were not significant at the 10% level in the original model (#3). Tables 5–7 summarize our results. In each table, all possible combinations of variable removals were tested, and the models were evaluated based on their likelihood functions. Note that in each table, all listed models have the same number of variables (since in each table, the same number of variables has been removed from the original model #3), so selecting the best model in each table based on the likelihood function is reasonable. The highlighted row in each table indicates the model that performed the best (i.e., had the lowest -2LL function).

Table 5: Performance of Choice Models with Three Variables Removed:

Model ID	Fwy Speed Variable Used?	On-Ramp Flow Variable Used?	Weather Variable Used?	Same-Interchange Control Ramp Used?	New Ave Control Ramp Used?	Atlantic Blvd Control Ramp Used?	Constant Used?	-2 Log Likelihood
38				✓	✓	✓	✓	2320.807
39			✓		✓	✓	✓	2321.477
40			✓	✓		✓	✓	2321.619
41			✓	✓	✓	✓		2320.949
42		✓			✓	✓	✓	2319.949
43		✓		✓		✓	✓	2319.908
44		✓		✓	✓	✓		2319.403
45		✓	✓			✓	✓	2320.623
46		✓	✓		✓	✓		2320.561
47		✓	✓	✓		✓		2321.411
48	✓				✓	✓	✓	2319.28
49	✓			✓		✓	✓	2319.505
50	✓			✓	✓	✓		2319.425
51	✓		✓			✓	✓	2320.194
52	✓		✓		✓	✓		2319.828
53	✓		✓	✓		✓		2319.843
54	✓	✓				✓	✓	2319.007
55	✓	✓			✓	✓		2318.467
56	✓	✓		✓		✓		2318.392
57	✓	✓	✓			✓		2319.044

Table 6: Performance of Choice Models with Four Variables Removed:

Model ID	Fwy Speed Variable Used?	On-Ramp Flow Variable Used?	Weather Variable Used?	Same-Interchange Control Ramp Used?	New Ave Control Ramp Used?	Atlantic Blvd Control Ramp Used?	Constant Used?	-2 Log Likelihood
60					✓	✓	✓	2321.483
61				✓		✓	✓	2321.639
62				✓	✓	✓		2321.001
63			✓			✓	✓	2322.2
64			✓		✓	✓		2322.044
65			✓	✓		✓		2323.036
66		✓				✓	✓	2320.625
67		✓			✓	✓		2321.018
68		✓		✓		✓		2321.674
69		✓	✓			✓		2323.749
70	✓					✓	✓	2320.194
71	✓				✓	✓		2319.834
72	✓			✓		✓		2319.889
73	✓		✓			✓		2320.331
74	✓	✓				✓		2319.046

Table 7: Performance of Choice Models with Five Variables Removed:

Model ID	Fwy Speed Variable Used?	On-Ramp Flow Variable Used?	Weather Variable Used?	Same-Interchange Control Ramp Used?	New Ave Control Ramp Used?	Atlantic Blvd Control Ramp Used?	Constant Used?	-2 Log Likelihood
77						✓	✓	2322.21
78					✓	✓		2322.291
79				✓		✓		2323.158
80			✓			✓		2325.068
81		✓				✓		2324.629
82	✓					✓		2320.346

Measuring the Risk of the best candidate choice models

Having now identified the best performing choice model (#3) among those with all seven classes of predictor variables, as well as the best models of those variants of Model #3 with low-significance variables removed (i.e., #15, #32, #56, #74, and #82), we use Leave-One-Out Cross Validation (LOOCV) to evaluate the predictive power of these choice models, quantified by the level of risk associated with each candidate model.

The results of LOOCV for choice models #3, 15, 32, 56, 74, and 82 are shown in Table 8. The aggregate risk-minimizing model is also included for comparison. Of the choice models, the one with the lowest risk is #74, with a risk of 0.0004478.

Table 8: Risk Measurements for Each of the Optimal Models

Model ID	Fwy Speed Variable Used?	On-Ramp Flow Variable Used?	Weather Variable Used?	Same-Interchange Control Ramp Used?	New Ave Control Ramp Used?	Atlantic Blvd Control Ramp Used?	Constant Used?	Average Risk
Agg.	✓		✓		✓		✓	0.000457
3	✓	✓	✓	✓	✓	✓	✓	0.001101
15	✓	✓		✓	✓	✓	✓	0.001019
32	✓	✓		✓	✓	✓		0.000797
56	✓	✓		✓		✓		0.000613
74	✓	✓				✓		0.000448
82	✓					✓		0.000481

Choice Model Approach for Yield Line at New Ave ramp

Choice Model Results for New Ave

Because the procedure used for developing a choice model for New Ave is analogous to the steps used for the San Gabriel Blvd on-ramp earlier, only the summary tables will be included, along with supplemental notes as needed. For explanations of the tables, refer to their analogs in the documentation for the San Gabriel Blvd on-ramp earlier. For example, Table 2-N is the New Ave counterpart of Table 2 from the San Gabriel Blvd section.

Table 1-N: Performance of Choice Models, Set 1

Model ID (for reference)	Speed Variable Used	On-ramp Flow Variable Used	-2 Log Likelihood
1	Freeway Speed	On-Ramp Flow	923.628
2	Freeway Speed	Natural Log of On-Ramp Flow	923.389
3	Freeway Speed	On-Ramp Flow above 60 v/5m	924.221
4	Freeway Speed	On-Ramp Flow above 70 v/5m	924.377
5	Freeway Speed	On-Ramp Flow above 80 v/5m	925.225
6	Fwy Speed below 55	On-Ramp Flow	924.772
7	Fwy Speed below 55	Natural Log of On-Ramp Flow	924.326
8	Fwy Speed below 55	On-Ramp Flow above 60 v/5m	2318.995
9	Fwy Speed below 55	On-Ramp Flow above 70 v/5m	2320.666
10	Fwy Speed below 55	On-Ramp Flow above 80 v/5m	2319.527

Table 2-N: Coefficient Details for Model #2

The coefficients have the following properties:

Variable	Coefficient	Standard Error	Significance
Natural Log of On-Ramp Flow	.110	.081	.174
Freeway Speed passing on-ramp	.023	.021	.259
Weather is sunny	-.258	.307	.400
Hourly yield rate at Same-Interchange Control Ramp	-4.197	1.487	.005
Hourly yield rate at Same-Side Control at San Gabriel Bl	-1.500	1.371	.274
Hourly yield rate at Same-Side Control Ramp at Atlantic Bl	1.510	2.348	.520
Constant	2.017	2.250	.370

Based on the level of significance for these coefficients, we will next test different versions of this model, each having one of the above insignificant (at a 10% level) variables removed. As was the case with the San Gabriel on-ramp, there are six variables that are not significant at a 10% level, meaning that we have the same number of combinations in Tables 3-N through 7-N as we did in Tables 3-7.

Note that the same-interchange control ramp was included in every model of Tables 3-N through 7-N, since it was found to be significant at a 10% level in our initially selected optimal model (i.e., Model #2).

Table 3-N: Performance of Choice Models with One Variable Removed:

Model ID	On-Ramp Flow Variable Used?	Fwy Speed Variable Used?	Weather Variable Used?	Same-Interchange Control Ramp Used?	San Gabriel Control Ramp Used?	Atlantic Blvd Control Ramp Used?	Constant Used?	-2 Log Likelihood
13		✓	✓	✓	✓	✓	✓	925.23
14	✓		✓	✓	✓	✓	✓	924.662
15	✓	✓		✓	✓	✓	✓	924.113
16	✓	✓	✓	✓		✓	✓	924.586
17	✓	✓	✓	✓	✓		✓	923.803
18	✓	✓	✓	✓	✓	✓		924.193

Table 4-N: Performance of Choice Models with Two Variables Removed:

Model ID	On-Ramp Flow Variable Used?	Fwy Speed Variable Used?	Weather Variable Used?	Same-Interchange Control Ramp Used?	San Gabriel Control Ramp Used?	Atlantic Blvd Control Ramp Used?	Constant Used?	-2 Log Likelihood
21			✓	✓	✓	✓	✓	926.359
22		✓		✓	✓	✓	✓	925.789
23		✓	✓	✓		✓	✓	926.633
24		✓	✓	✓	✓		✓	925.464
25		✓	✓	✓	✓	✓		926.541
26	✓			✓	✓	✓	✓	925.329
27	✓		✓	✓		✓	✓	926.809
28	✓		✓	✓	✓		✓	924.994
29	✓		✓	✓	✓	✓		929.756
30	✓	✓		✓		✓	✓	925.03
31	✓	✓		✓	✓		✓	924.667
32	✓	✓		✓	✓	✓		924.737
33	✓	✓	✓	✓			✓	925.492
34	✓	✓	✓	✓		✓		924.66
35	✓	✓	✓	✓	✓			925.704

Table 5-N: Performance of Choice Models with Three Variables Removed:

Model ID	On-Ramp Flow Variable Used?	Fwy Speed Variable Used?	Weather Variable Used?	Same-Interchange Control Ramp Used?	San Gabriel Control Ramp Used?	Atlantic Blvd Control Ramp Used?	Constant Used?	-2 Log Likelihood
38				✓	✓	✓	✓	926.885
39			✓	✓		✓	✓	928.727
40			✓	✓	✓		✓	926.541
41			✓	✓	✓	✓		932.869
42		✓		✓		✓	✓	926.928
43		✓		✓	✓		✓	926.128
44		✓		✓	✓	✓		926.885
45		✓	✓	✓			✓	927.282
46		✓	✓	✓		✓		926.903
47		✓	✓	✓	✓			927.902
48	✓			✓		✓	✓	927.103
49	✓			✓	✓		✓	925.786
50	✓			✓	✓	✓		929.88
51	✓		✓	✓			✓	927.752
52	✓		✓	✓		✓		929.762
53	✓		✓	✓	✓			935.644
54	✓	✓		✓			✓	926.035
55	✓	✓		✓		✓		925.094
56	✓	✓		✓	✓			926.361
57	✓	✓	✓	✓				926.067

Table 6-N: Performance of Choice Models with Four Variables Removed:

Model ID	On-Ramp Flow Variable Used?	Fwy Speed Variable Used?	Weather Variable Used?	Same-Interchange Control Ramp Used?	San Gabriel Control Ramp Used?	Atlantic Blvd Control Ramp Used?	Constant Used?	-2 Log Likelihood
60				✓		✓	✓	928.918
61				✓	✓		✓	927.156
62				✓	✓	✓		932.895
63			✓	✓			✓	929.412
64			✓	✓		✓		932.873
65			✓	✓	✓			938.743
66		✓		✓			✓	927.657

67		✓		✓		✓		927.174
68		✓		✓	✓			928.346
69		✓	✓	✓				928.183
70	✓			✓			✓	928.129
71	✓			✓		✓		929.883
72	✓			✓	✓			935.758
73	✓		✓	✓				935.8
74	✓	✓		✓				926.616

Table 7-N: Performance of Choice Models with Five Variables Removed:

Model ID	On-Ramp Flow Variable Used?	Fwy Speed Variable Used?	Weather Variable Used?	Same-Interchange Control Ramp Used?	San Gabriel Control Ramp Used?	Atlantic Blvd Control Ramp Used?	Constant Used?	-2 Log Likelihood
77				✓			✓	929.668
78				✓		✓		932.901
79				✓	✓			938.765
80			✓	✓				939.026
81		✓		✓				928.551
82	✓			✓				935.938

Table 8-N: Risk Measurements for Each of the Optimal Models

Model ID	On-Ramp Flow Variable Used?	Fwy Speed Variable Used?	Weather Variable Used?	Same-Interchange Control Ramp Used?	San Gabriel Control Ramp Used?	Atlantic Blvd Control Ramp Used?	Constant Used?	Average Risk
Agg.	✓		✓	✓		✓	✓	0.000447
2	✓	✓	✓	✓	✓	✓	✓	0.006400
17	✓	✓	✓	✓	✓		✓	0.006271
34	✓	✓	✓	✓		✓		0.006102
55	✓	✓		✓		✓		0.005859
74	✓	✓		✓				0.005802
81		✓		✓				0.005028

Of the choice models, the one with the lowest risk is #81, with a risk of 0.005028. Since this is an order of magnitude greater than the risk associated with the aggregate model, **this suggests that even our best choice model does not perform better than the aggregate risk-minimizing model from before.**