Heckle and Chide: Results of a randomized road safety intervention in Kenya*

James Habyarimana† and William Jack‡

July 27, 2010

Abstract

In economies with weak enforcement of traffic regulations, drivers who adopt excessively risky behavior impose externalities on other vehicles, and on their own passengers. In light of the difficulties of correcting *inter*-vehicle externalities associated with weak third-party enforcement, this paper evaluates an intervention that aims instead to correct the *intra*-vehicle externality between a driver and his passengers, who face a collective action problem when deciding whether to exert social pressure on the driver if their safety is compromised. We report the results of a field experiment aimed at inducing passengers to exert social pressure on their drivers. Evocative messages encouraging passengers to speak up were placed inside a random sample of over 1,000 long-distance Kenyan minibuses, or matatus, serving both as a focal point for, and to reduce the cost of, passenger action. Independent insurance claims data were collected for the treatment group and a control group before and after the intervention. Our results indicate that insurance claims fell by a half to two-thirds, from an annual rate of about 10 percent without the intervention, and that claims involving injury or death fell by 60%. While we cannot definitively disentangle the mechanisms through which this intervention works, results of a driver survey eight months into the intervention suggest passenger heckling was a contributing factor to the improvement in safety.

*We gratefully acknowledge the financial support of the Center for Global Development and the Safaricom Foundation, and thank Channa Commanday and Bright Oywaya of ASIRT-Kenya, the Kenyan branch of the Association for Safe International Road Travel, an international NGO. We thank Mr. Tom Gichuhi of the Association of Kenyan Insurers, senior executive officers of four large Kenyan insurance companies, and executive officers of the 21 *matatu* savings and credit cooperatives who assisted us in this project. We also thank the editor and two anonymous referees, as well as Nada Eissa, David Evans, Luca Flabbi, Garance Genicot, Vijaya Ramachandran, Roger Lagunoff and Tavneet Suri for discussions, and seminar participants at Georgetown, the World Bank, and the Kenya Medical Research Institute. We also acknowledge the pro-bono contributions of George Wanjohi and Saracen Media in Nairobi, and John Wali and volunteers from Junior Achievement Kenya. We thank Lauren Marra for excellent research assistance. Finally we thank Philomena Wanjiru, David Gitahi, Asman Wesonga and Nadeem Karmali for their tireless and professional work in leading our team of 20 field workers in implementing the study. All errors are our own.

†Georgetown University, Public Policy Institute, e-mail: jph35@georgetown.edu

‡Georgetown University, Department of Economics, e-mail: wgj@georgetown.edu

1 Introduction

This paper reports the results of a field experiment aimed at empowering individuals to exert pressure on service providers. The specific context is that of long-distance road transportation services in Kenya, where it is popularly believed that otherwise rational young males are transformed, Jekyll-and-Hyde-like, into irrational death-seekers when they occupy the driver's seat of a minibus, or *matatu*. Our intervention motivates passengers to exercise their power as consumers, literally giving them a voice, by encouraging them to speak up, to heckle and chide the driver when his behavior compromises their safety.

Individuals can be empowered to help themselves either by providing them with resources that tip the balance of economic power in their favor, or by changing the decision-making environment in which they operate in a way that increases their bargaining power or political clout. Although both types of empowerment will likely be resisted - either by those who fund the resource transfers or by those who see their own economic and/or political power eroded - they are often seen as potentially powerful development initiatives that enhance not only the economic well-being of the poor, but their human dignity as well (World Bank, 2004). However, some recent studies have examined the extent to which the poor exercise the power conferred upon them: just as leading a horse to water is not enough to make it drink, mandating empowerment (Banerjee et. al., 2008) might not induce the beneficiaries to seize control of their destinies. The intervention in this paper motivates individuals to do just that.

Much of the recent literature on beneficiary empowerment has focused on its role in improving the delivery of public services, particularly in health and education (see Bjorkman and Svensson (2008), Svensson and Reinnika (2006) and Olken(2007) on roads), many of which are free or highly subsidized at the point of use. By contrast, this paper investigates the impact of consumer empowerment on the delivery of a privately provided service that people pay for as and when they use it - long-distance road transportation. In Kenya, large buses and smaller, 14-seater minivans, known locally as *matatus*, are the primary mode of long distance transportation.[1] Our study focuses on the quality, in particular the safety, of long distance *matatu* travel. For a variety of reasons addressed below, the price mechanism might not be effective in ensuring efficient quality in this market.

Long distance transportation services in much of the developing world are provided by the private sector and account for a significant share of road traffic injuries and fatalities, which in turn constitute a large and increasing share of both deaths and the disease burden in the developing world. The World Health Organization (2004) reported that 1.2 million people died from road traffic injuries in 2002, 90% in low- and middle-income countries, about the same number as die of malaria. In addition, between 20 and 50 million people are estimated to be injured or disabled each year. Road traffic accidents constitute the largest share, 23%, of deaths due to injury, nearly twice as many as the 14% due to war and violence combined. Traffic accidents were ranked as the 10th leading cause of death in 2001, and are projected to be the third or fourth most important contributor to the global disease burden in 2030 (Lopez *et. al.* 2006). By that date, road accidents are projected to account for 3.7 percent of deaths worldwide - twice the projected share due to malaria (Mathers and Loncar, 2006).[2] In addition to the gains to operators and insurance and health service providers, given that the primary consumers of these services are prime-age adults, reducing the extent of road traffic injuries and fatalities could confer large welfare gains on households (see Mohanan (2008), Beegle et. al. (2008) and Evans and Miguel (2007)).[3]

Many interventions to reduce road accidents have been undertaken in developed economies, including programs to reduce the volume of driving, to improve the safety features of road

[1] In the early days of 14-seater bus service, the fare for the most typical ride was three (*tatu* in Kiswahili) Kenyan Shillings.

[2] Country level data are generally less reliable. Odero *et al.* (2003) suggest that fatality rates in Kenya are extremely high with 7 deaths from 35 road crashes every day, and that the impact of prevailing interventions is dismal. According to a Ministry of Health Report, in 1996 traffic accidents were the third leading cause of death after malaria and HIV/AIDS (Government of Kenya, 1996). More recent estimates suggest that over 3,000 individuals died in road traffic related incidents in 2008 (Association of Kenyan Insurers, 2008)

[3] Road accidents affect the elite as well as the poor. Recent example include the death in March 2009 of the wife of Zimbabwe's prime minister, Morgan Tsvangirai, the serious injury of then future Kenyan president Mwai Kibaki during the election campaign of 2002, and the involvement of former Kenyan president Daniel arap Moi in a serious road accident in 2006.

networks, and to enforce traffic regulations more effectively.[4] Publicity campaigns have focused on educating road users, and some, most notably in Australia and New Zealand, have employed shock therapy to get their message across. For example, an adverstising campaign in New Zealand aimed at reducing speeding and drunk-driving, and encouraging the use of safety belts, was found to have an impact on road deaths (Guria and Leung, 2004). Fewer studies of interventions in developing countries exist and while the estimated results of these studies are not causal, measured effects are large. The introduction of speed bumps at certain accident hot-spots in Ghana was associated with a 35% reduction in accidents and a 55% reduction in fatalities (Afukaar et al., 2003). Bishai et al. (2008) found that higher intensity police patrols were associated with a 17% reduction in accident rates in Uganda. Perhaps more creatively, in Bogotá, Colombia, mimes were used to ridicule pedestrians and drivers who flaunted traffic rules.[5]

In our field experiment, we randomize an intervention aimed at empowering *matatu* passengers to exert pressure on drivers to drive more safely. The intervention was simple and cheap: stickers with evocative messages intended to motivate passengers to take demonstrative action - to heckle and chide a dangerous driver - were placed in just over half of 2,276 recruited *matatus*. High rates of compliance were ensured by running a monthly lottery among drivers of participating treatment *matatus*, who could win up to 5,000 Kenyan Shillings (about $60, or roughly one week's wages) if their vehicle was found to have all stickers intact upon inspection by our field staff. Our main outcome data were collected independently from four insurance companies that together cover more than 90% of these vehicles, and who were unaware of our intervention at the time it took place. We use insurance claims data for treatment and control vehicles in the two year window bracketing the insertion of the stickers. We identify an impact on driver behavior that is both statistically significant and economically large. Our intention-to-treat estimate indicates that the stick-

[4] A comprehensive review of such interventions can be found in World Health Organization (2004), Chapter 4.

[5] This intervention, supported by the Mayor of Bogotá, Antanas Mockus, was not rigorously evaluated, but reportedly enjoyed high levels of popularity (Caballero, 2004).

ers are associated with a reduction in insurance claims rates of about a half, from a projected counterfactual annual baseline claims rate of about 10 percent. Our instrumental variables estimate of the average treatment effect on the treated yields an even higher estimate of the impact among compliers.

Furthermore, we find that this result is largely due to a reduction in claim events where the driver was at fault. We also document a large reduction in claims involving injury or death. We present some suggestive evidence that this effect is associated with consumer empowerment and action from surveys of both drivers and passengers. In particular, drivers of treated vehicles report significantly more passenger complaints than drivers of control *matatus*. While our results are imprecise, conditional on experiencing a risky trip, passengers in treatment *matatus* are more likely to express concerns to their driver.

In general, disentangling the mechanisms through which stickers alter driver behavior is difficult. Drivers might be responding to greater heckling in treatment vehicles, but they might also be anticipating such heckling and pre-empting it by driving more safely. Alternatively, drivers might be responding to the stickers themselves directly.[6] Direct driver effects could operate through the stickers changing the beliefs of drivers about accidents or their beliefs about the preferences of the owner. In all these cases, our identification of the impact of the stickers on driver behavior remains valid. However, if the results are due to *ex post* sorting by the drivers between treatment and control vehicles, then the results would not be attributable to any change in driving behavior. We argue below that such sorting is very unlikely, and cannot be large enough to account for a significant part of the effect in any case. Similarly, we argue that the monthly lottery does not account for the measured effects.

Economists typically deem bad driving to be inefficient because of the externality it imposes on other drivers. Regulation of such behavior by a third party, such as the police, can correct this market failure, but if the police are corrupt and themselves difficult to

[6] Even though the stickers are placed behind the driver, he is no doubt aware of their existence.

monitor, a speeding fine can be as much an opportunity for extortion and a source of rents as it is a Pigouvian tax. Other sources of inefficiency might also exist, not between vehicles, but within vehicles. Bargains between driver and passenger might be difficult to enforce, and could lead drivers to actively extort "safety" money from passengers. Alternatively, a positive externality could exist between one passenger and another, if a costly individual intervention has benefits for all, in terms of safer driving. Our intervention is motivated by these two sources of inefficiency. We believe our stickers encourage passengers to exert social pressure on the driver, literally heckling him to take account of the costs that his actions impose on them. This could be either because they reduce the individual psychic costs of heckling, or because they provide a focal point upon which passengers acting non-cooperatively can nonetheless coordinate their actions.

Social pressure has been observed or advocated in a variety of other settings. Microfinance institutions have relied on it to improve loan repayment rates and profitability, by making self-selected, and hence relatively homogeneous, groups liable for loans.[7] Similarly, in the political domain, Gerber *et al.* (2008) find that the prospect of disclosure of (non-) participation to an individual's household and neighbors, which they interpret as a form of social pressure, leads to higher voter turn-out.[8] In the field of public health, McGuckin *et al.* (2001, 2004) report results from an intervention similar to ours in which patients were motivated to ask their doctors if they had washed or sanitized their hands. That intervention shares the feature of empowering consumers to question authority with ours, albeit in a vastly different context, with positive effects on provider performance.

In other contexts economists have succeeded in estimating non-zero effects of social pressure, most notably on the response of European football referees to home crowd biases (Dawson and Dobson, 2008, Garicano *et al.*, 2005). Although these careful studies iden-

[7] The empirical evidence in support of this contractual design is however mixed (Armedariz de Aghion and Morduch, 2000, Morduch, 1998, Pitt, 1999), and some MFIs have recently moved away from the strategy.

[8] A growing literature on collective action and ethnic diversity suggests that social pressure is relatively more effective within groups than between groups (see for example Khwaja (2008), Miguel and Gugerty (2005), Okten and Osili (2004) and Bardhan (2000)).

tify statistically significant impacts of social pressure on referee behavior, as measured for example by the length of injury time granted, they do not appear to be large enough to have economically meaningful consequences, in terms of affecting the identity of winners and losers. Within the environment of a 14-seater *matatu*, social pressure exerted by passengers on the driver is arguably more benign than that exerted by football crowds on referees, and issues of favoritism and lack of fairness, which are the focus of much of that literature, are turned on their heads. Indeed, our intervention is aimed at giving voice to passengers in order for them to more effectively exert the social pressure that is a corrupting influence in other settings. In this context, social pressure is generated in a way that produces economically large and socially important beneficial effects.

The rest of the paper is organized as follows. Section 2 describes the context, data and empirical strategy. We present the results of the intervention and discuss possible mechanisms in section 3, and conclude in section 4.

2 Context and experimental design

In this section we describe the salient features of the long distance *matatu* sector, and the environment in which driver and passenger actions are taken, to further motivate our intervention. We then describe the intervention in detail and review the extent to which our experimental design was implemented in practice.

2.1 The *matatu* sector

There are about 50,000 *matatus* operating in Kenya, providing both intra-city transportation in Nairobi, Mombasa, Kisumu and other large urban areas, as well as inter-city services across much of the country. *Matatu* ownership is broad, with many owners having fleets of just a handful of vehicles. Those plying the inter-city routes are organized into either Savings And Credit Co-Operatives (SACCOs), or limited liability companies, which range in size from 20-30 to around 500 vehicles. These SACCOs and companies engage in scheduling and other organizational activities associated with the provision of *matatu* services, and provide financial services to both owners and drivers. In our sample, about 70 percent of drivers operate a single *matatu* on a long-term basis, while the others are either temporary drivers, or rotate across vehicles within a particular SACCO.

Road travel options are differentiated by both price and some observable vehicle characteristics, including the number and comfort of seats. Within the 14-seater *matatu* sector, quality differences are potentially associated with reputations of particular SACCOs, reflecting marketing policies, driver recruitment and training, vehicle maintenance, etc. Drivers are officially paid a fixed daily wage and owners are responsible for the running and maintenance costs of the *matatu.* A small fraction of owners were former drivers.

The effectiveness of this intervention will depend in part on who consumes long distance services, how frequently they use these services, and their experience of road traffic accidents. At the outset of the study, we surveyed passengers who had just completed an inter-city trip by *matatu*, and found that more than half had made a similar long-distance trip in the last week, and 80 percent had done so in the last month. Furthermore, *matatu* users are predominantly of prime working age, with two thirds of the respondents between the ages of 20 and 40. One third of the respondents reported feeling that their life was in danger on a *matatu* trip in the previous month but half of the respondents had never experienced a life-threatening event. Heterogeneity in passenger experiences underlines the importance of a potential mechanism of the intervention we evaluate: increasing the salience of risky driver behavior and coordinating passenger action.

2.2 Driver and passenger behavior

A *matatu* driver acts as an agent of both the vehicle's owner and its passengers. In the absence of any agency problems, and assuming a well-functioning market for transportation services, we would expect the quality of such services - as defined by speed, safety, convenience, comfort, etc. - to be efficient, reflecting the marginal costs and benefits of improved quality. In particular, passengers would get the safety they pay for.

However, the relationships between drivers and both owners and passengers are fraught with agency problems, in which case it might be difficult for either party to reliably purchase safe driving. From the passenger's perspective, once on board s/he is, quite literally, captive and cannot expect to recoup her/his monetary outlay if disatisfied with the service. In addition, the market is sufficiently thick and anonymous that it is difficult for a given driver to establish and maintain a reputation for good driving.[9]

From the perspective of owners, information on actual driver behavior is virtually impossible to observe, so rewards for careful driving are infeasible. Outcome variables upon which performance incentives might be conditioned – such as crashes or officially recorded traffic violations – are characterized by low signal-to-noise ratios. Making the driver the residual claimant in terms of liability for damage would expose these workers to excessive risk, while conditioning wages on police reports of bad driving would likely provide yet another opportunity for corruption.

On the other hand in practice, it appears that drivers *are* residual claimants with regard to marginal fare collections: if anything, this could increase the incentives of drivers to drive recklessly, if it would mean reaching a potential passenger ahead of other *matatus*. Finally, under Kenyan law, all public service vehicles are required to have third party insurance, which further attenuates incentives for safe driving.

Self preservation arguably provides the strongest incentive for safe driving, although the behavior exhibited by some drivers suggests it is not always operative.[10] In any case, the fact that *matatus* are used by a broad range of Kenyan society, across which incomes, and hence the value of life, vary significantly, suggests that for at least some trips the driver's optimal point on the risk-speed frontier will not reflect the preferences of his passengers.

[9] In addition, there are no fixed schedules that would enable passengers with private information about driver quality to choose when to travel.

[10] An explanation consistent with these facts is excessive optimism about the likelihood and severity of accidents (see for example Lovallo and Kahneman (2003) and Camerer and Lovallo (1999)).

Paying the driver to slow down (or indeed, to speed up) is unlikely to be observed, due both to free-rider problems among the passengers, and to the incentives the driver would face for outright extortion. Instead, we suggest that passengers can affect driver behavior through social pressure: by adopting a "heckle and chide" strategy.

2.3 Experimental design

Our empirical strategy compares outcomes of *matatus* in which stickers had been inserted with those of *matatus* without such stickers. In our pre-recruitment survey we presented passengers with a variety of stickers and asked which would be more likely to induce them or others to voice complaints directed to the driver in the event of poor or dangerous driving. Three types of messages were presented to respondents: the first set had text-only messages (in both English and Kiswahili, the national language), in which individuals were encouraged to take action; the second group of stickers included similar text messages, but with supporting images with a "soft-touch"[11]; the third group represented fear stimuli, in which forceful messages about the consequences of accidents were accompanied by explicit and gruesome images of severed body parts.

The results of the pre-intervention survey (not reported in detail here) indicated support for the effectiveness of both the fear stimuli and simple text messages, but not for the soft-touch approach.[12] The chosen stickers are shown in Figure 1 in the Appendix. Stickers were placed on the metal panel between a passenger window and the ceiling of the vehicle, ensuring that at least one sticker was within the eye view of each passenger sitting in the main cabin. The stickers were not placed in direct view of the driver or the passengers in the front cabin.

FIGURE 1 GOES HERE

[11] This category included subtle visual information such as a missing parent at a baptism or graduation.

[12] In future work we hope to be able to evaluate the differential impact of these alternative interventions, but due to sample size constraints, the intervention we adopted in this study was a combination of the five most effective stickers.

Although recruitment was at the individual driver level, we first sought cooperation from the SACCOs operating long-distance *matatu* services in Kenya, and obtained a letter from the management expressing support for our project. The major towns among which our sampled *matatus* operated are illustrated in Figure 2. In all, 21 SACCOs agreed to participate, and just three refused.[13] At the initial recruitment, participating SACCOs provided us with lists of license plates of vehicles in their fleets.

FIGURE 2 GOES HERE

In light of our pilot experience, which revealed that vehicle lists were of variable quality, and during which non-participation rates were observed to be reasonably low, we simplified the recruitment protocol and adopted a field-based sampling procedure. Under this strategy, if a *matatu* had been recruited at the pilot stage, it was again recruited and its randomly assigned treatment/control status was maintained. Each additional observed *matatu* from a participating SACCO was eligible for recruitment, and assignment to the treatment group based on the final numeric digit of its license plate (odd = treatment, even = control).

In addition, a follow-up survey undertaken soon after the pilot recruitment period found very low rates of sticker retention among treatment vehicles. To address this problem, at the second recruitment phase we implemented a weekly lottery that was to run throughout the remaining study period. All complying treatment vehicles were eligible for the lotteries, and three randomly chosen winners were inspected by our field staff.[14] If an inspected *matatu* was found to have retained all five stickers, the driver would receive a monetary prize: first prize was 5,000KSh (about $US60), second prize was 3,000KSh ($US35), and third prize was 2,000KSh ($US25).[15]

[13] SACCO non-participation reflected the extent to which officials could act on behalf of a large group of owners.

[14] At recruitment, we requested drivers provide us with their cell phone numbers, or a number at which they could be reached. To increase the perceived expected winnings, the treatment group was divided into 5 groups of roughly 200 *matatus* each. Each group's lottery was run every 5 weeks.

[15] Implementing the lottery was challenging, particularly given security concerns in and around the bus stations. The winning license plate numbers were randomly drawn off-site, after which one of our field staff would contact the driver and inspect the vehicle. If it was found to be in compliance, another field

The structure of the project and its objectives, were explained to each driver, as was the voluntary nature of his participation in the study.[16] Each driver in the treatment or control groups was asked to sign an informed consent form. Those selected to be in the treatment group were asked to accept all five stickers, although compliance with this request was incomplete (see below).

Tables 1 and 2 report descriptive statistics of vehicles and drivers respectively, for the treatment and control groups by random assignment. These data suggest that the randomization performed well, there being only one observable variable exhibiting a statistically significant difference between the two groups. This one source of difference between treatment and control groups could however be quite important, as it is the share of drivers who reported having had an accident in the last 12 months (second last row in Table 1). Among those assigned to treatment the self-reported accident rate was 1.5%, while among the assigned control group the rate was just 0.4% (p-value < 0.05). However, when we examine the insurance claims data, this difference disappears (see last row), suggesting that responses to this question may have been affected by treatment status. Indeed, drivers were administered the recruitment questionnaire *after* they were assigned to the treatment or control group, and those in the treatment group may have been induced to think more about their accident experiences, or even to exaggerate them. In any case, we do not use driver reports as our main outcome variable.

TABLE 1 GOES HERE

TABLE 2 GOES HERE

Compliance to the randomized assignment was high but not perfect. Table 3 reports that about 84 percent of vehicles assigned to the control group complied, and that the same share of those assigned to the treatment group took at least one sticker, with 68.5 percent taking all five, and 8.0 percent taking just three (typically the three text-only stickers).

staff member would be informed by phone, and would send money via M-PESA, a cell-phone based money transfer system, to the driver. The driver would confirm on the spot receipt of the prize.

[16] Our field staff encountered no female drivers, although a number of SACCO executives are women.

TABLE 3 GOES HERE

Imperfect compliance to the randomized assignments, either due to driver self-selection or fieldworker error, yielded some statistically significant differences in characteristics by actual assignment, as reflected in panel A of Table 4. However, the difference in self-reported accident rates that was significant for true assignment was narrower for actual assignment (the rates were 1.3% and 0.6% respectively). It is conceivable, although not necessarily obvious, that this small narrowing of the difference in self-reported accidents reflects selective adoption of the treatment (even by some assigned to the control) by relatively safe drivers, which would bias the OLS results in favor of finding an effect. To avoid any such bias, our estimates of the average treatment effect rely on intent-to-treat and instrumental variable estimation strategies.

TABLE 4 GOES HERE

2.4 Data and empirical strategy

In addition to baseline data collected at recruitment, we were granted access to a comprehensive database of claims data from four insurance companies that cover over 90 percent of long-distance *matatus* in our sample (see panel B in Table 4). There are three possible concerns associated with the use of insurance claims data as an outcome measure. Firstly, as claims are filed by drivers, owners, or passengers there is likely measurement error in observed accidents. However, we do not believe that the decision to file a claim is systematically correlated with randomized assignment to treatment since insurance companies were unaware of which vehicles were participating in the study and the owners and drivers were unaware of the source of our outcome data. While the resulting classical measurement error has implications for precision, it should not bias our results. The second issue is that we do not have access to insurance claims data on the entire sample of vehicles in the study. Panel B of Table 4 compares selected vehicle, trip and driver characteristics on the basis of whether we have claims data or not. There are no differences in vehicle and trip characteristics across these two groups, but we observe two significant differences in the driver characteristics between them: drivers of *matatus* for which we have claims data are more likely to have secondary schooling and are less likely to operate a single vehicle. It is difficult to say whether these differences impinge on the representativeness of the claims sample. Finally, we do not observe whether or how soon a vehicle involved in the claim continues to operate after the claim-generating event. Our simplifying assumption that each *matatu* continues to operate after an accident biases the result against us finding an effect of the intervention.

The claims data were collected for the period January 2006 through May 2009. We use annualized insurance claims rates as an outcome measure, as well as evidence based on our own coding of the description of the accidents such as whether the driver was at fault, and whether injuries or fatalities occurred.

Our accident-related data were complemented by surveys of both passengers and drivers of treatment and control *matatus*, fielded in November 2008, about eight months after recruitment. These surveys elicited information about experiences on the most recent trip, and on trips taken during the previous week and month. Reports by both passengers and drivers of the frequency of heckling, and by passengers of the safety of trips, are used as outcome variables to explore the mechanisms by which the stickers may impact behaviors.

We are interested in estimating the average causal effect of the sticker intervention on the outcomes outlined above. Using outcome information before and after sticker insertion we estimate the following specification:

$$Y_{it} = \alpha + \beta_1 P_{it} + \beta_2 TR_i + \beta_3 P_{it} * TR_i + \beta_4 X_{it} + \eta_i + \varepsilon_{it} \tag{1}$$

where Y_{it} represents the annualized claim rate for *matatu* i during period t, P_{it} is an indicator that takes on the value of 1 for all time periods after recruitment and 0 otherwise, and TR_i is an indicator equal to 1 if the *matatu* was ‘treated’ and 0 otherwise. Finally X_{it} represents a set of covariates that might include the vehicle condition, and driver and route characteristics, and η_i represents unobserved fixed characteristics of the driver, route and vehicle.

The main parameter of interest is β_3 which captures the *net* change in the outcome variable Y_{it} for treated vehicles compared with those in the control group. A negative and significant coefficient indicates a statistically significant decline in the claims rates among treatment *matatus*. This estimate, and the alternatives described below, likely represent lower bounds on the true value of the parameter due to potential spillovers across treatment and control *matatus*. If the empowerment effect of the stickers on individual passengers is durable, those who have been exposed to the treatment may be induced to heckle their driver in future trips, even when traveling in control *matatus*.[17]

Employing ordinary least squares, the identifying assumption in order to recover the average treatment effect β_3 is

$$Cov(TR_i, \eta_i + \varepsilon_{it}) = 0 \tag{2}$$

That is, we require that unobserved factors captured by $\eta_i + \varepsilon_{it}$ are uncorrelated with treatment status. We implement two identification strategies and a number of robustness checks to establish the validity of our estimates.

Firstly since Tables 1 and 2 confirm the plausibility of the identifying assumption 2 when TR_i corresponds to the random assignment rule, we estimate the intent-to-treat parameter $\widehat{\beta}_3^{itt}$.

Secondly, as Table 3 demonstrates, compliance to assignment is not perfect:16% of *matatus* in the control arm did not comply with their assignment. In addition, just over 68% of *matatus* assigned to the treatment arm accepted all five stickers, and 16% of them accepted none. Since imperfect and possibly selective compliance can dilute the estimated effects of assignment to treatment, we use an instrumental variables strategy to estimate the average treatment effect on the treated, using the indicator for random assignment to the treatment group as an instrument for receiving the stickers. The resulting estimator, $\widehat{\beta}_3^{att}$,

[17]Similarly, drivers who move between treated and untreated matatus could be a source of spillovers.

represents the local average treatment effect of the stickers for the group of vehicles whose treatment status is affected by random assignment. In the results section below we present both estimators.

Both of these estimators assume parallel trends in claims rates across both the treatment and control groups. While random assignment should assure this, we present a number of robustness checks that allow for differential trends below.

3 Results

3.1 Effects on insurance claims

A visual summary of the results is presented in Figure 3, in which the trajectories of claims events per 1,000 *matatus* are shown, separately for vehicles assigned to treatment and control, from the first quarter of 2006 to the second quarter of 2009. The horizontal axis in Figure 3 measures calendar time. Given the considerable lag of 3-6 months in claim reporting, our data for the first two quarters of 2009 are incomplete. The vertical line indicates when recruitment of *matatus* started.[18] Not surprisingly, quarterly claims rates are very noisy, so that before recruitment we observe moderate albeit insignificant differences across the treatment and control groups. While the sign of the differences in claims rates between treatment and control vehicles oscillates before recruitment, it is consistently negative in the post recruitment period. In particular, claims rates for *matatus* assigned to receive the stickers are considerably lower in the quarters after recruitment (except for Q2 2009, which is very incomplete).

FIGURES 3 GOES HERE

Consistent estimation of β_3 requires that the trend in accident rates in the absence of treatment be parallel across both groups. As the fitted lines in the figure show, a parallel trends assumption is consistent with the pattern of observed outcomes across the two

[18] Recruitment started in February 2008 and was completed in the second quarter of 2008.

groups.[19] The claims rate trajectory for the control *matatus* starts a little lower but is marginally steeper so that the trend lines cross. We show below that our results are robust to allowing for group specific trends.

One concern suggested by Figure 3 is the apparent increase in accident rates for control *matatus* in the first three quarters of 2008. This apparent deviation from trend occurred immediately after the wide-spread violence that followed the December 2007 election. During this recovery, traffic volumes increased, particularly for long-distance trucks serving land-locked neighbors. Using comprehensive data from one of the insurance companies, we are able to calculate claims rates for larger 30-41 seater vehicles that, like our sample, predominantly provide long distance services along similar routes. We combine data on the number of active insurance policies (which serves as our denominator) with data on the claims for each quarter. The results are shown in Figure 4. Consistent with the results in Figure 3, quarters 2 and 3 of 2008 show higher levels of claims rates for these larger vehicles, suggesting a secular trend in accident rates over this period.

FIGURE 4 GOES HERE

Table 5 presents regression results that correspond to the two identification strategies discussed above: panel A presents the intent-to-treat estimates while panel B presents the effect of treatment-on-the-treated estimates. Columns (1) and (2) report the effect of the intervention on all claims. Columns (3) and (4) examine the effect of the intervention on claims where the driver is at fault and columns (5) and (6) examine the effect on claims that involve a physical injury or fatality. The even numbered columns include a fixed effect for the management cooperative (SACCO).

TABLE 5 GOES HERE

[19] In estimating the trend for the control matatus, we use information from 2006-2008 to determine the trend and avoid the bias inherent in using the incomplete information from 2009. The trend for the treatment group uses all the quarters in 2006 and 2007.

The intent-to-treat estimate of the impact of the intervention on all claims is a large negative and significant reduction in claims rates of five percentage points (p-value < 0.01). Including controls for the management group increases the magnitude of this effect only slightly. In 312 of the 362 claims events in our data (about 86 percent) that we could classify,[20] the *matatu* driver is recorded as being at fault. Using these data, columns (3) and (4) of panel A in Table 5 presents an ITT estimate of the impact of our intervention on driver-at-fault claims. The point estimate of -4.63 percentage points is statistically significant (p-value < 0.01) and represents a 53% reduction in driver-at-fault claims below the projected base. Finally, the intervention we evaluate appears to be particularly effective in reducing serious accidents. Our data include 227 claims with at least one injury or death. Using this as an outcome variable in columns (5) and (6) of Table 5, we estimate a 4 percentage point reduction in the rate of injury/death claims that is statistically significant (p-value < 0.01) and large, representing a 60% reduction in such accidents from the projected base of 6.65 percent.

In panel B, we present the results of an instrumental variables strategy in which we use random assignment as an instrument for receiving at least one sticker. The binary instrument produces a Wald estimator which scales the intent-to-treat estimates by the inverse of the difference in compliance rates between vehicles assigned to treatment and control, respectively. Given the results observed in Table 3, our average treatment-on-the-treated estimates are about 50% larger than the ITT estimates. In columns (1) and (2) we estimate a 7.3 percentage point reduction in all claims rates amongst those vehicles that are affected by the instrument. Relative to the projected claims rate, the LATE estimator suggests a decline in the rate of accidents of as much as 73% associated with the treatment. We observe even larger treatment effects for driver-at-fault and injury-death claims. While the usefulness of some IV results is legitimately questioned in the face of weak instruments and impact heterogeneity, we believe our strong first stage and high compliance rates make

[20] Two claims had no accompanying descriptions that could be used for this coding exercise.

this a credible estimate of the impact of the stickers. Among compliers to the instrument, around three-quarters of the accidents that would otherwise have occurred are avoided.

3.2 Robustness of Estimated Treatment Effect

In this sub-section we present a robustness test for the main results outlined above. Difference-in-differences estimates rely on the assumption of parrallel trends in the outcomes in the absence of the intervention. If this assumption does not hold, the measured treatment effect is biased due to trend differences between the groups. While random assignment should make such trend differences unlikely and Figure 3 suggests similar trends, nevertheless we include controls for group specific trends to specification 1. In particular we estimate the following specification at the quarterly level.

$$y_{it} = \gamma_0 + \gamma_1 Post_{it} + \gamma_2 TR_i + \gamma_3 Post_{it} * TR_i + \phi_m + (\lambda + \delta * TR_i)t + \eta_i + \varepsilon_{it} \qquad (3)$$

where y_{it} is an indicator for whether *matatu* i had an accident claim in quarter t. $Post$ and TR are defined as before in specification 1. ϕ_m represents calendar quarter fixed effects and controls for seasonal variation in accident rates. The key additional control is an allowance for group specific trends captured by λ and δ.

In Table 6, we present both linear probability (panel A) and marginal probit (panel B) estimation of the specification above. The results in column(1) in Table 6 present analogous results to column (1), panel A of Table 5 with the sole difference that the results in Table 5 represent annualized claim rates. In columns (2) and (3) we include a secular time trend and calendar quarter fixed effects. Both of these additions should not change the estimated impact of the intervention since the timing of recruitment is orthogonal to the assignment. In column (4) we include a group specific trend which allows the pre-recruitment trends in accident rates to differ across the two groups. The point estimates remain virtually unchanged and indicate large and significant reductions in quarterly claims rates of about 1.2 percentage points (panel A) and about 1 percentage point (panel B) respectively. In the case of the linear probability model of panel A, this represents a reduction of about 50 percent compared with the projected rate.

The result in Column (4) is the regression equivalent of a tripple difference-in-difference estimator in which a pre-recruitment Diff-in-diff is substracted from the estimates in Table 5. In a working paper version of this paper.we show the the pre-recruitment diff-in-diff is positive and insignificant consistent with the trend lines in Figure 3.

TABLE 6 GOES HERE

Both of these results suggest that the main results presented above are unlikely the consequence of trend differences in the two groups.

3.3 Cost-effectiveness measures

Our intervention achieves significant reductions in insurance claims and accidents at a very low cost. We compare this cost to that of other health interventions by performing a basic cost-effectiveness analysis in which we calculate the cost per year of life saved. In our data, 11 percent of claims involved at least one death, although we do not know the actual number of deaths associated with each such accident. In our baseline case, the projected claims rate in the treatment group is about 10 percent, which the treatment reduces by five percentage points. Assuming the same rate of reduction in accidents involving a death as in accidents involving injuries or death (we are better able to esitmate the impact on the latter), the intervention thus reduced the number of accidents including a death by about 6.0 per year per thousand vehicles. Conservatively we assume an average of two fatalities per accident including a death.[21] We further make a rough assumption that the average death results in a loss of 20 years of life. The intervention thus saved about 1,200 years of life.

[21] Newspaper headlines reporting "Another 14 people killed in matatu crash" are not uncommon.

The cost of the intervention was roughly \$2 per vehicle for the stickers, and \$5 per vehicle per year for the lottery, or a total of \$7,000 per 1,000 vehicles per year. The cost per life year saved is thus about \$5.80 (0.8% of per capita GDP) including the lottery costs, and \$1.70 (0.2% of per capita GDP) counting only the material costs. Measures of the cost per Disability Adjusted Life Year (DALY) gained would be smaller still once the reduced numbers of injuries were included. The cost-effectiveness of the intervention, including lottery costs, is thus lower than that of childhood vaccination, which at \$7 per DALY gained is considered to be among the most cost-effective health interventions available, and it is an order of magnitude lower than virtually all interventions that are considered to be "good buys" in development, such as tuberculosis therapy using the directly observed treatment – short course (DOTS) strategy (\$102 per DALY), and improved emergency obstetric care (\$127 per DALY), as calculated by Jamison et al. (2006), and reported in Disease Control Priority Project (2008, Figure 2).

4 Evidence on mechanisms of change

In this section we present some suggestive evidence for potential mechanisms that underlie the reduction in accident rates estimated above. As we indicated in the introduction, we do not have the data to definitively discriminate amongst all plausible mechanisms underlying our results. However, we present two pieces of evidence in support of passenger action mechanisms and discuss the plausibility of a number of other mechanisms including direct effects on drivers, *ex post* sorting of drivers, and the effects of the lottery.

4.1 Survey evidence

One of the mechanisms discussed in the introduction is that stickers empower passengers traveling in *matatus* to voice their concerns over bad driving and that the resulting social pressure conditions the behavior of the driver. To investigate if this could be the cause of the observed differential in claims rates identified above, we analyze data from a survey fielded in November 2008 of drivers, plus up to three of their passengers. A total of 284 vehicles were sampled for this survey.[22]

We face two difficulties in detecting evidence for this mechanism. First, even if the stickers are effective in empowering passengers, we might observe little or no difference in heckling if drivers of treatment vehicles quickly learn to adapt their behavior to minimize passenger complaints. On the other hand, whether heckling is observed in equilibrium or not, we might expect passengers to report their trips as being safer in treatment matatus. Secondly, given the rarity of traffic accidents, events that generate heckling will also be rare. Compounding this power problem is the fact that, despite the weekly lottery, after 8 months a considerable number of the treatment vehicles had lost some or all of their stickers. Table 7 shows that, among our sample of 284 *matatus*, the share with all five stickers had fallen from 44% at recruitment to 18% eight months later, and the share with at least one sticker had fallen from 53% to 37%.

TABLE 7 GOES HERE

Table 8 presents evidence of heckling from the survey of drivers (panel A) and passengers (panel B) and passenger-reported safety ratings (panel C). We present intent-to-treat estimates for all outcome measures. Note that this considerably limits our ability to find any evidence for this mechanism as a result of low sticker retention.

The results are suggestive of passenger heckling as one of contributors to the reduction in accident rates. In rows (1) and (2) of panel A, we estimate the effect of assignment on the likelihood that the driver reports passenger heckling in the past week and most recent trip. The point estimate in row (1) has the right sign but is imprecisely estimated. The sign of the coefficient in row (2) is wrong but again imprecise. However, in OLS results not reported here, we find substantial and marginally significant effects of having a sticker eight months

[22] We interviewed 306 drivers, but 22 of them were operating vehicles that had not been recruited earlier.

into the study. In particular, drivers of vehicles with stickers at the time of the survey were about three times more likely to report passenger heckling.[23]

TABLE 8 GOES HERE

We next turn to self-reported evidence of passenger action in panel B of Table 8. Sampled passengers were asked to report if they or any other passengers had said something to the driver/conductor about reckless driving behavior on the just concluded trip. In order to avoid conflating potentially frivolous actions with legitimate heckling, we control for the reported safety of the trip. In particular, passengers were asked to rank the safety of the trip on a scale of 1 to 10, with 1 denoting no danger and 10 denoting life-threatening. While nearly 45% of the respondents reported that they "could not say", we define a trip to be reported as safe if the passenger reports a rating equal to or less than 5. For our current purposes we create an indicator for whether at least one passenger had rated the safety of the trip as dangerous (a rating of 6 or higher). Evidence for the passenger heckling mechanism is then captured by the extent to which there is a greater likelihood of heckling on trips deemed dangerous by at least one passenger. We present ITT estimates for four different outcomes that correspond to the rows in panel B of Table 8: likelihood of heckling by (1) the respondent (2) any passenger (3) at least two respondents (4) all respondents. The latter two outcomes represent a crude measure of the extent to which the intervention facilitates collective passenger action and the unit of observation is the vehicle.[24] The coefficient of interest is the interaction of the indicator for stickers and whether at least one passenger rated the trip as unsafe.

[23]In a simple OLS estimation of the effect of stickers on heckling, non-random removal or depreciation of stickers could bias our results. On the one hand, dangerous drivers might have removed them, either in advance or in response to unwelcome heckling as they learned about their effectiveness over time. This would work against finding evidence of passenger action in treated vehicles. On the other hand, if the stickers provided drivers who otherwise lacked self-control with an effective enforcement technology, removal could be concentrated in the pool of relatively safe drivers, who simply find them distasteful and perhaps bad for business. This would bias our results in favor of finding an effect. Although we cannot distinguish econometrically between these two directions of bias, we believe the former is more plausible and highly likely to dominate the latter.

[24]For two or more respondent reports of heckling we are unable to condition on the same dangerous event.

Our estimates for this parameter are of the wrong sign in rows (1) - (3), but in all cases are very imprecise. In row (4) that estimates the likelihood that all correspondents heckle the driver, we obtain the right sign but once again the coefficient is statistically insignificant.

One way in which this mechanism could operate is by making passenger heckling a credible threat to reckless driving. In the absence of more objective measures of driving behavior, we rely on passenger ratings of safety of the just concluded trip. Our results in panel C report the results of an ordered probit estimation across three safety ratings categories (safe trip, cannot say, unsafe). About two thirds of all passengers in the control *matatus* rated the most recent trip as safe according to this definition. The ordered probit estimate in panel C has the right sign but is very imprecisely estimated.

While the evidence above suggests that passenger action may well lie at the heart of the observed effects, we cannot definitively rule out a number of other potential mechanisms. For instance while passenger ratings of safety do not confirm this (see panel C of Table 8), it is possible that a driver's beliefs regarding the preferences of the vehicle's owner, over either passenger safety or the life of the vehicle, could be affected by this intervention. More direct observations of driver behavior might shed more light on the plausibility of this mechanism.

4.2 Driver sorting

Alternatively, although the *ex ante* assignment of stickers to drivers was random, the *ex post* assignment may have exhibited sorting. That is, it is possible that rather than stickers having altered the behind-the-wheel behavior of drivers, either directly or via passenger action, they induced sorting of drivers across treatment and control *matatus*. For example, suppose reckless drivers in treated vehicles tended to switch to control vehicles, or to exit this labor market entirely, while safe drivers in control vehicles on average moved to treated *matatus*. Such sorting behavior could have led to the observed changes in claims rates, but would not have been associated with any change in driving practices per se. We present three pieces of evidence suggesting that this kind of *ex post* sorting does not constitute a likely explanation of the results.

First, the share of treatment vehicles within each *matatu* cooperative (SACCO) is about half, so sorting within SACCOs is definitely feasible. However, the authority to hire and fire drivers rests not with the SACCO, but with the owners of the vehicles. But since *matatu* ownership is very diffuse, sorting within an individual owner's fleet (which can be as small as one or two vehicles) is unlikely to generate our measured effects. And given the costs of sorting out of treament vehicles, it would be much easier for the drivers to remove the stickers than to find an eligible and willing partner with whom to switch.

Second, it is possible that this sorting operates more on the participation margin, if reckless drivers tend to quit the treatment group. Data on driver tenure suggests that the median tenure is about 10 months and that while overall turnover since recruitment has been high (an average of 39%), there is no statistically significant difference in turnover rates across treatment and control vehicles (41 vs 37%). This holds true amongst the drivers assigned to a single vehicle.

And third, selective sorting could take place within just those SACCOs that have a policy of regularly rotating drivers across vehicles, as long as such rotation was non-random. However, our results could be driven by selective sorting among the relatively small group of drivers in such SACCOs only if there was a high concentration of claims among "reckless" drivers. The insurance claims data from the period before our intervention do not support this pattern. Although the identity of the driver is not recorded in the data, we do know that before our intervention, fewer than 8% of all claims were associated with multiple-claim vehicles (and possibly drivers). Overall, these three pieces of evidence suggest that while we cannot rule out driver sorting as a response to the intervention, the scale at which such sorting could be occurring cannot explain the results obtained above.

4.3 Direct effects of the lottery

Finally, we discuss the possibility that the presence of the lottery, designed to improve sticker retention, could itself lead to our empirical results. Recall that drivers who accepted all five stickers at recruitment were divided into 5 groups of roughly 200 vehicles, and that each week on a five-week rotating basis, members of one of the groups were eligible to win one of three prizes if, when randomly drawn, upon inspection they were found to have retained all five stickers. The total prize money each week of 10,000 shillings (about 2 weeks' wages) was awarded in three amounts (5,000, 3,000 and 2,000) to three different winners.

The lottery itself could have changed the beliefs of drivers of treatment vehicles about the likelihood and consequences of an accident. Alternatively, while the rules of the lottery were very explicit, and drivers were told that eligibility was based on *sticker retention* and not an accident-free record, it is still conceivable that drivers with stickers might have misconstrued the lottery as as reward for safe driving. The policy implications of such findings would, of course, be radically different to those that would otherwise be drawn.

On the first point, knowledge of the lottery and its association with the road safety project were not confined to treatment vehicles alone or lottery nominees. Inspection of stickers was done at parking lots where control and treatment drivers interacted quite frequently, and where awareness of the role of the sticker inspector was clear to both. As a result, we believe that any small differences in road safety salience attributable to the lottery across the two groups is unlikely to explain the large effect measured above.

On the second point, which is potentially of greater concern, the payment is likely to have been too small to alter driving behavior. Expected winnings were very low (equal to wages equivalent to about 20 minutes work), and even if drivers had unreasonable priors of winning, the first prize was considerably less than what a driver could make by squeezing in one extra trip (unreported to owner) per month.

Nonetheless, to address this second issue more quantitatively, we investigate the beliefs that drivers would have had to maintain in order that the observed reduction in claims rates could be rationalized in terms of a response to the misguided belief that safe driving would increase the chance of winning the lottery. This kind of exercise is of course laden with assumptions and can only inform the analysis if the results suggest wildly counterfactual driver beliefs. In fact, we find that such extreme beliefs, plus an impossibly high response of accidents to speed reductions, are indeed necessary to support the claim that the lottery was the driving force behind the impact we observe.

The key parameter in this exercise is the elasticity of accidents with respect to speed, estimates of which are not available in Kenya or other developing countries to our knowledge. Ashenfelter and Greenstone (2009) report data for the US suggesting an elasticity of fatalities of about 4, which provides a benchmark against which to compare our data.[25] As we illustrate below,[26] even if a driver (i) thought he would win the lottery *with certainty* (instead of with

---

[25] The approximately equal estimated proportional impacts of the intervention on all claims, claims in which the driver was at fault, and claims involving an injury or death suggest that this fatality elasticity is a good proxy for the elasticity of all accidents.

[26] Each week three prizes totaling 10,000 shillings were awarded. We assume driver risk neutrality and denote the size of the average weekly prize by $x = 10,000/3$. The probability the driver assesses to winning a prize, conditional on not having had an accident, is denoted p, and the expected winning each week are px. Let w be the driver's weekly wage, and denote $z = x/w$ as the ratio of the average prize to the wage. In order to reduce the chance of being involved in an accident, thereby increasing his chance of winning, the (misinformed) driver slows down. We want to compare the expected increase in winnings to the cost this would impose on him.

Let π_0 be the weekly probability of having an accident under the assumption of no behaior change. (The projected counterfactual annual accident rate among treated matatus during the year following the intervention was approximately 10 percent, so $\pi_0 = 0.1/52.$). Drivers in the treatment group experienced a claims rate about half the projected rate. Assuming for the sake or simplicity a constant proportional reduction over the year, their actual weekly probability of having an accident was $\pi_1 = \pi_0/2$, which is also the change in the probability, $\Delta\pi$. Engaging in this behavior change increases expected weekly winnings by $B = \Delta\pi px = \Delta\pi pzw$.

The expected cost per week of slowing down is the wage times the extra time taken, $w\Delta t$, which is approximately equal to

$$C = w(\Delta s/s)$$

where s is the average speed of the vehicle.. Define the elasticity of accidents, a, with respect to speed, s, by

$$\varepsilon = \left(\frac{\Delta a}{a}\right) \Big/ \left(\frac{\Delta s}{s}\right).$$

The relationship between speed and accident rates is not known in Kenya. Ashenfelter and Greenstone (2009) present fatality and speed data from the US that suggests an elasticity of fatalities with respect to speed of about 4. (In their data, a 4.55% reduction in speed is associated with a 15.46% reduction fatalities.) Thus the cost incurred by the driver in reducing accidents by this much is approximately

$$C = \frac{w}{\alpha\varepsilon},$$

average weekly probability 0.003), (ii) was sure of reducing his chance of an accident *to zero*, and (iii) thought that there was a *single prize of 10,000 shillings* every week, the elasticity of accidents with respect to speed would still need to be more than *30 times* larger that the US estimate for the expected financial benefit of slowing down to outweigh the expected costs. In light of the evidence, recently reviewed by Delavande et al. (2009), that people in developing countries generally understand the concept of probability, we believe this calculation, while clearly subject to wide margins of error, nonetheless strongly suggests the lottery itself did not affect driver behavior enough to account for any meaningful share of the estimated effects of the intervention.

5 Conclusions

We have presented evidence that a very cheap intervention can alter the behavior of drivers in the context of long distance minibus transportation services in Kenya. Our estimates consistently suggest that the intervention reduced the number of incidents leading to an insurance claim by about a half. The intervention empowers passengers to question the authority of the driver when his behavior endangers their lives. Our evidence suggests that by voicing their concerns in a coordinated fashion, passengers exert social pressure that is effective in discouraging dangerous driving.

Although the size of the effect that we estimate is very large, we argue that it is nonetheless plausible. Our intervention is neither intended to raise the ambient noise level in a *matatu*, inducing its passengers to constantly heckle and interfere with the driver, nor to

where $1/\alpha = \Delta a/a \approx 1/2$. This cost is less than the expected benefit, $C < B$, if

$$1 < \alpha\varepsilon\Delta\pi pz.$$

Using our data the right hand side of this expression is approximately $2\times4\times\left(\frac{0.05}{52}\right)\times\left(\frac{3}{1{,}000}\right)\times\left(\frac{2}{3}\right) = 1/65,000$. That is, for a driver to respond only to the incentive of a lottery whose eligibility criteria he misinterpreted, and not to the stickers or the response they evoked on the part of passengers, he would need to over-estimate the right hand side of the inequality condition above by a factor of 65,000. Even if he thought he would win the lottery with certainty ($p = 1$), was sure of reducing his chance of an accident to zero ($\Delta p = \frac{0.1}{52}$) and thought that there was a single prize of 10,000 shillings ($z = 2$), the elasticity of accidents with respect to speed would still need to be 32.5 times larger that the US estimate for the condition above to be satisfied.

create a generally hostile environment inside the vehicle. Instead, it allows individuals to overcome their inhibitions against voicing their fears in the moments before a potential crash. These events, while tragically common, are still statistically rare, suggesting that, even in the absence of dynamic effects whereby drivers pre-empt heckling by driving more safely, heckling will be observed relatively infrequently. Our evidence that driver-reported heckling rates are three times higher in *matatus* with stickers is consistent with our findings of a large reduction in the claims rate.

Our results represent a step towards identifying the kinds of interventions that can tip the balance of power in favor of consumers when the price mechanism is not fully effective in guaranteeing quality service provision. This is not simply a redistribution of bargaining power however, but a mechanism which allows a small group of consumers to better coordinate their actions to ensure they get what they have paid for.

Leveraging passenger action represents a promising intervention to address a rising problem of road traffic injury and fatalities in similar settings in other developing countries. In the context of private provision among a diffuse set of owners and weak enforcement from the police, there are no alternatives that have the cost and informational advantages of passenger-based interventions. However, while we have identified such an intervention that is effective, further research is required to determine the optimal design of this approach. As in other arenas, the size of the treatment effect is potentially sensitive to the types of information and framing used (see Bertrand et. al. (2007) and Saez (2009)). Understanding which content and framing strategies are most effective in mobilizing passenger action is chief among these questions. In future work, we hope to estimate the relative impact of evocative messages compared to simple imperatives common in public health campaigns. This would have obvious implications for other information dissemination programs such as anti-smoking, safe sex, and immunization campaigns. A second aim of future research would be to gain a clear understanding of the mechanisms, the duration over which such interventions are effective, and how frequently messages need to be updated.

References

Afukaar, F.K., P. Antwi, and S. Ofosu-Amah (2003): "Pattern of road traffic injuries in Ghana: implications for control," *Injury Control and Safety Promotion*, 10:69-76.

Ashenfelter, Orley and Michael Greenstone (2004): "Using Mandated Speed Limits to Measure the Value of a Statistical Life." *Journal of Political Economy*, 112(1).

Banerjee, Abhijit., Rukmini Banerji, Esther Duflo, Rachel Glennerster, and Stuti Khemani (2008). "Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India," mimeo

Beegle, Kathleen, De Weerdt, Joachim and Dercon, Stefan (2008). "Adult Mortality and Consumption Growth in the Age of HIV/AIDS," *Economic Development and Cultural Change*, vol. 56

Bertrand, Marianne., Karlan Dean., Mullainathan Sendhil., Shafir, Eldar., and Zinman, Jonathan (2005). "What is Psychology Worth? A Field Experiment in the Consumer Credit Market" National Bureau of Economic Research Working Paper 11892.

Bishai, D. Asiimwe, B, Abbas, S, Hyder, A. Bazeyo, W. (2008): "Cost Effectiveness of Traffic Enforcement: Case Study From Uganda Injury Prevention," *Injury Prevention*, 14:223-227.

Bjorkman, Martina and Svensson, Jakob, (2008). "Power to the people : evidence from a randomized field experiment of a community-based monitoring project in Uganda," *forthcoming Quarterly Journal of Economics*

Dawson, Peter and Stephen Dobson (2008): "The Influence of Social Pressure and Nationality on Individual Decisions: Evidence from the Behaviour of Referees," IASE/NAASE Working Paper Series, Paper No. 08-09.

Disease Control Priorities Project (2008): "Using cost-effectiveness analysis for setting health priorities," March, http://www.dcp2.org/file/150/DCPP-CostEffectiveness.pdf

Delavande, Adeline, Xavier Giné and David McKenzie (2009): "Measuring subjective expectations in developing countries: a critical review and new evidence," BREAD working paper.

Evans, David and Edward Miguel (2007): "Orphans and schooling: A longitudinal analysis," *Demography* 44(1): 35-57

Garicano, Luis, Ignacio Palacios-Huerta, and Canice Prendergast (2005): "Favoritism Under Social Pressure," *The Review of Economics and Statistics*, 87(2): 208-216.

Gerber, Alan, Donald Green and Christopher W. Larimer (2008): "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment, *American Political Science Review*, 102(1): 33-48.

Government of Kenya (1996): Ministry of Health. Health Information System, 1996 Report. Nairobi. Government Printers.

Guria, Jagadish and Joanne Leung (2004): "An evaluation of a supplementary road safety package," *Accident Analysis and Prevention*, 36(5), 893-904.

Jamison, D.T., J.G. Breman, A. R. Measham, G. Alleyne, M. Claeson, D.B. Evans, P. Jha, A. Mills, and P. Musgrove (2006): *Disease Control Priorities in Developing Countries, 2nd Edition*, New York, Oxford University Press.

Lopez, A., Mathers, C., Ezzati, M., Jamison, D., and Murray, C. (2006): "Global and Regional Burden of Disease and Risk Factors, 2001: Systematic Analysis of Population Health Data". *Lancet* 367: 1747-1757.

Mathers, C. and Loncar, D. (2006): "Projections of Global Mortality and Burden of Disease from 2002 to 2030," *PloS Medicine* 3(11): 2011-2030.

McGuckin M, Waterman R, Storr J, Bowler CJW, Ashby M, Topley K, Porten L. (2001): "Evaluation of Patient Empowering Hand Hygiene Programme in UK," *The Journal of Hospital Infection*, 48:222-227.

McGuckin M, Taylor A, Martin V, Porten L, Salcido R. (2004): "Evaluation of a Patient Education Model for Increasing Hand Hygiene Compliance in an In-Patient Rehabilitation Unit," *American Journal Infection Control*,32:235-8

Miguel, Edward and Mary Kay Gugerty, (2005), "Ethnic Diversity, Social Sanctions, and Public Goods in Kenya," *Journal of Public Economics* 89, pp. 2325-2368

Mohanan, Manoj (2008). "Consumption Smoothing and Household Responses: Evidence from Random Exogenous Health Shocks" mimeo Harvard University.

Odero W., Khayesi, M., and Heda, P.M., (2003). "Road Traffic Injuries in Kenya: Magnitude, Causes and Status of Intervention". Injury Control and Safety Promotion 10 1-2 53-61.

Olken, Ben (2007). "Monitoring Corruption: Evidence from a Field Experiment in Indonesia". *Journal of Political Economy* 115 (2).

Saez, Emmanuel (2009). "Details Matter: The Impact of Presentation and Information on the Take-up of Financial Incentives for Retirement Savings" *American Economic Journal: Economic Policy* 1 204-228.

Svensson, J. and Reinnika, R. (2005). "Fighting Corruption to Improve Schooling: Evidence from a Newspaper Campaign in Uganda." *Journal of.the European Economic. Association.* 3 (April–May): 259–67

World Bank (2004): *World Development Report 2004: Making Services Work for People*, Oxford, OUP.

World Health Organization (2004): *World Report on Road Traffic Injury Prevention*, WHO, Geneva.

World Health Organization: http://rbm.who.int/wmr2005/html/exsummary_en.htm

Appendix

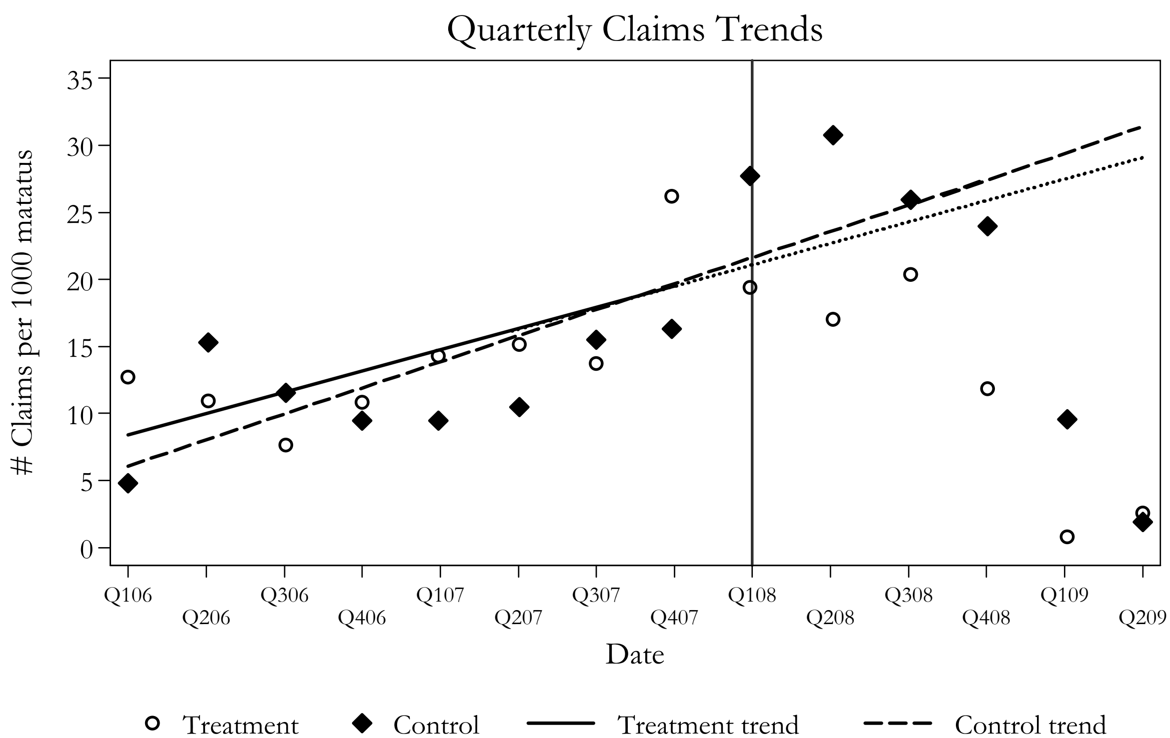
Figure 1: Stickers Inserted in Treated *Matatus*

Hey! If he drives badly, will you arrive?
STAY AWAKE. BE ALERT. SPEAK UP!

Hey! Will you complain after he causes an accident?
STAY AWAKE. BE ALERT. SPEAK UP!

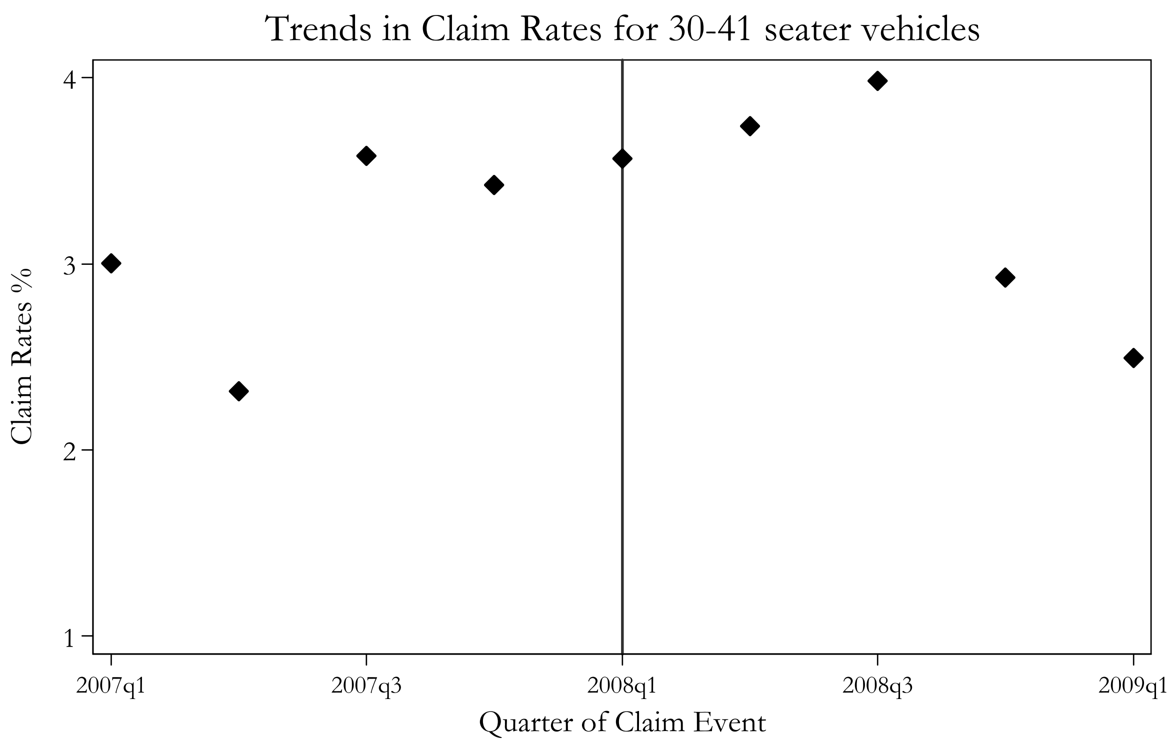
Figure 2: Major towns served by sampled long-distance *matatus.*

Figure 3: Quarterly Claims Trends For Recruited Sample: January 2006 – May 2009



Note: The figure presents the number of insurance claims by quarter between January 1 2006 and May 25 2009. *All* insurance claims are used to construct this figure. Solid and dashed lines represent fitted linear trends for the treatment and control group. Trend lines are linear best fits using data from 2006-2008 for the control *matatus* and 2006-2007 for the treatment *matatus.* We make the simplifying assumptions that *matatus* continue to operate after a claim event and were operating throughout this period.

Figure 4: Claims Rates for 30-41 Seater Vehicles that provide Long Distance Transportation Services



Note: Figure generated using data on the ratio of the number of 30-41 seater vehicles reporting a claim in a particular quarter to the total number of insured 30-41 seater vehicles in that quarter. Data is drawn from one of the larger insurance companies for which seat class is provided. The vertical line corresponds to the quarter in which the majority of *matatus* were recruited.

Table 1: Vehicle Characteristics

Vehicle characteristic	Control	Treatment	Difference Significant	Total
Odometer reading	356506.85	361386.98	No	359111.75
	(7236.26)	(6350.53)		(4781.66)
	[327266]	[343603]		[336454]
Seating Capacity	14.52	14.52	No	14.52
	(0.05)	(0.05)		(0.03)
Proportion use tout	0.45	0.48	No	0.47
	(0.02)	(0.01)		(0.01)
Number of weekly trips	20.19	19.60	No	19.88
	(0.36)	(0.30)		(0.23)
Average daily distance, kilometers	420.48	414.10	No	417.07
	(6.14)	(5.33)		(4.04)
	[400]	[400]		[400]
Proportion with an installed speed governor	1.00	1.00	No	1.00
	(0.001)	(0.001)		(0.001)
Share owned by large Cooperative	0.49	0.51	No	0.50
	(0.02)	(0.01)		(0.01)
Involved in accident in last 12 months, self reported	0.004	0.015	Yes	0.01
	(0.002)	(0.004)		(0.002)
Insurance claim filed in last 12 months before recruitment	0.061	0.071	No	0.055
	(.008)	(.007)		(.005)
Number of observations	1006	1155		2161

Notes: Standard errors in ( ); Medians in [ ]. The table presents mean/median of vehicle characteristics by *treatment assignment*. The sample is restricted to *matatus* for which information on random assignment is available. 115 *matatus* that could not be matched to the initial assignment list are dropped.

Table 2: Driver Characteristics

Driver Characteristic	Control	Treatment	Difference significant	Total
Has access to phone[a]	0.96	0.98	No	0.97
	(0.01)	(0.00)		(0.00)
Owns a phone[a]	0.89	0.91	No	0.90
	(0.01)	(0.01)		(0.01)
% less than 30 years old	18.5	16.2	No	17.3
	(3.4)	(3.0)		(2.3)
% 30-40 years old	54.8	56.1	No	55.5
	(4.3)	(4.1)		(3.0)
% Primary schooling	22.8	26.2	No	24.6
	(3.5)	(3.5)		(2.5)
% Secondary schooling	13.9	14.7	No	14.3
	(2.8)	(2.8)		(2.0)
% Married	74.8	77.0	No	76.0
	(3.7)	(3.5)		(2.5)
Number of children	2.0	2.0	No	2.0
	(0.1)	(0.1)		(0.1)
Proportion drivers assigned	0.72	0.70	No	0.71
to one car only	(0.04)	(0.04)		(0.03)
Proportion drivers started	0.37	0.41	No	0.39
after recruitment	(0.04)	(0.04)		(0.03)
Median driver tenure, days	296	305.5		304
Number of observations	139	145		284

Notes: Standard errors in ( ); Medians in [ ]. The table presents mean/median of driver characteristics by *treatment assignment*.
[a] Statistics reported in these rows are based on the sample of all recruited *matatus*. The statistics reported in the rest of the table are based on a random sample of 284 *matatu* drivers who were surveyed about 6 months after recruitment.

Table 3: Compliance to the Intervention

Number of stickers actually inserted	True assignment (%)	
	Treatment	Control
0	16.1	84.4
1	3.6	0.3
2	3.1	0.2
3	8.0	0.5
4	0.7	0.1
5	68.5	14.5
Total	100.0	100.0

Notes: The table presents the number of intervention stickers inserted at recruitment by *treatment assignment*. The sample is restricted to *matatus* for which information on random assignment is available. 115 *matatus* that could not be matched to the initial assignment list are dropped.

Table 4: Selection in and out of treatment

Covariates	Control	Treatment	Difference Significant (5%)
PANEL A: Selection on Observables			
Has access to phone*	0.96	0.99	Yes
	(0.01)	(0.00)	
Owns a phone*	0.87	0.93	Yes
	(0.01)	(0.01)	
Odometer reading	354580.98	363246.75	No
	(7092.53)	(6461.47)	
	[324568]	[346064]	
Seating Capacity	14.56	14.48	No
	(0.05)	(0.05)	
Proportion use tout	0.44	0.49	Yes
	(0.02)	(0.01)	
Number of weekly trips	20.00	19.76	No
	(0.36)	(0.30)	
Average daily distance, kilometers	418.65	415.63	No
	(5.74)	(5.67)	
	[400]	[400]	
Proportion with an installed speed governor	1.00	1.00	No
	(0.001)	(0.002)	
Share owned by large Cooperative	0.47	0.53	Yes
	(0.02)	(0.01)	
Involved in accident in last 12 months, self reported	0.006	0.013	Yes
	(0.002)	(0.003)	
Insurance claim filed in last 12 months before recruitment	0.062	0.07	No
	(.008)	(.007)	
Number of observations	1035	1126	
PANEL B: Selection on Outcome Data	No Claims Data	Claims Data	
Share of vehicles (%)	8.8	91.2	
Vehicle and Trip Characteristics			
Odometer reading	280422.67	322448.77	No
	(45015.03)	(12620.3)	
	[199992.5]	[292980]	
Seating Capacity	14.56	14.32	No
	(.25)	(.11)	
Number of weekly trips	19.12	18.68	No
	(1.37)	(.52)	
Average daily distance, kilometers	433.48	422.39	No
	(39.25)	(11.14)	
Share owned by large Cooperative	0.64	0.48	No
	(0.1)	(.03)	
Driver Characteristics			
Owns a phone*	0.88	0.88	No
	(.07)	(.02)	
% Secondary schooling	5.28	15.19	Yes
	(3.95)	(2.13)	
% Married	88.0	74.81	No
	(6.63)	(2.71)	
Proportion drivers assigned to one car only	0.84	0.69	Yes
	(.07)	(.03)	

Notes: Standard errors in ( ); Medians in [ ]. Panel A presents mean/median of driver/vehicle characteristics by *actual treatment status*. The sample is restricted to *matatus* for which information on random assignment is available. 115 *matatus* that could not be matched to the initial assignment list are dropped. Panel B presents selected means by *whether the insurer of a particular vehicle was one of the four firms that provided claims data.*

Table 5: Regression Results

	All Claims		Driver-at-fault claims		Injury/Death Claims	
Panel A: Intent-to-Treat	(1)	(2)	(3)	(4)	(5)	(6)
Post	0.029	0.030	0.025	0.026	0.018	0.018
	(0.013)*	(0.012)*	(0.011)*	(0.011)*	(0.009)+	(0.009)*
Assigned to Treatment	0.010	0.009	0.011	0.011	0.011	0.011
	(0.010)	(0.011)	(0.010)	(0.010)	(0.008)	(0.008)
PostXAssigned to Treatment	**-0.050**	**-0.051**	**-0.046**	**-0.047**	**-0.040**	**-0.041**
	(0.016)**	**(0.016)****	**(0.014)****	**(0.014)****	**(0.012)****	**(0.012)****
Constant	0.061	0.042	0.052	0.039	0.038	0.036
	(0.008)**	(0.013)**	(0.007)**	(0.012)**	(0.006)**	(0.010)**
Panel B: IV Estimates						
Effect of Treatment on the Treated	**-0.073**	**-0.075**	**-0.068**	**-0.069**	**-0.059**	**-0.060**
	(0.023)**	**(0.023)****	**(0.021)****	**(0.021)****	**(0.017)****	**(0.017)****
Controls for SACCO		X		X		X
Observations	4322	4318	4322	4318	4322	4318
R-squared	0.003	0.02	0.002	0.01	0.002	0.01

Notes: Robust standard errors in parentheses. * significant at 5%; ** significant at 1%. Table reports the estimates of ordinary least squares regression in specifications (1-4) and instrumental variables estimates in specifications (5-6). The dependent variable is the annualized rate of a claim-generating accident for each *matatu* in the sample. We make a simplifying assumption that *matatus* continue to operate after a claim event and were operating throughout the pre- and post-recruitment period. First stage *F*-stat reports the *F*-stat of the test of the null that random assignment to treatment does not predict actual treatment status at recruitment. The sample excludes 3% of recruited vehicles for which treatment assignment information could not be reliably established.

Table 6: Testing Robustness of ITT estimates: Group Specific Time Trends

	Dependent Variable: Indicator for Claim-backed accident in Quarter			
Panel A: LPM	(1)	(2)	(3)	(4)
Post	0.011	0.006	0.006	0.006
	(0.003)**	(0.004)	(0.005)	(0.006)
Assigned to Treatment	0.002	0.002	0.002	0.001
	(0.002)	(0.002)	(0.002)	(0.004)
PostXAssigned to Treatment	**-0.012**	**-0.012**	**-0.012**	**-0.013**
	(0.004)**	**(0.004)****	**(0.004)****	**(0.006)***
Time trend		0.001	0.001	0.001
		(0.000)*	(0.000)	(0.001)
Assigned to Treatment *time trend				0.000
				(0.001)
Constant	0.013	0.009	0.008	0.008
	(0.001)**	(0.002)**	(0.003)*	(0.003)*
Panel B: Marginal Probit Estimates.				
PostXAssigned to Treatment	**-0.009**	**-0.009**	**-0.009**	**-0.008**
	(0.002)**	**(0.002)****	**(0.002)****	**(0.003)***
Calendar quarter fixed effects			X	X
Observations	28783	28783	28783	28783
R-squared	0.00	0.00	0.00	0.00

Notes: Standard errors clustered at the vehicle level in ( ); * significant at 5%; ** significant at 1%. Panel A presents the results of a linear probability estimation of the likelihood of having an accident by quarter. Panel B presents marginal probit estimates for each specification show above. This estimation uses all insurance claims matched to the experimental sample from January 2006 to May 2009. The sample is restricted to *matatus* for which information on random assignment is available. 115 *matatus* that could not be matched to the initial assignment list are dropped. We make a simplifying assumption that *matatus* continue to operate after a claim event and were operating throughout this period.

Table 7: Sticker Retention

Number of stickers in vehicle	Distribution at Recruitment (%)	Distribution in November 2008 (%)
	(1)	(2)
0	46.5	63.0
1	2.1	4.9
2	2.8	4.2
3	4.2	7.4
4	0.3	2.5
5	44.0	18.0
Total	100.0	100.0

Notes: Table reports the distribution of stickers for the random sample of *matatus* surveyed 8 months after recruitment. Column (1) reports the distribution at recruitment while column (2) reports the distribution 8 months after recruitment.

Table 8: Evidence on Passenger Action Mechanisms

Dependent Variable	Assigned to Treatment	Unsafe Trip	Assigned to Treatments* Unsafe Trip	Number of observations
Panel A: Driver Reports of Heckling				
(1) Driver reports heckling (Past week)	0.027 (0.034)	-	-	259
(2) Driver reports heckling (Last trip)	-0.027 (0.027)	-	-	259
Panel B: Passenger Reports (Most recent trip)				
(1) Any passenger expressed concern	-0.005 (0.088)	0.172 (0.070)*	-0.022 (0.097)	788
(2) Respondent expressed concern	0.014 (0.071)	0.084 (0.064)	-0.043 (0.078)	788
(3) At least two respondents expressed concern	0.092 (0.130)	0.300 (0.101)**	-0.092 (0.145)	260
(4) All three respondents expressed concern	-0.058 (0.079)	0.081 (0.077)	0.031 (0.093)	260
Panel C: Passenger Perceptions of Safety (Most recent trip)				
(1) Safety rating	-0.007 (0.078)	-	-	788

Notes: Robust standard errors in (). * Significant at 5%, ** significant at 1% level. Panel A reports the results of a linear probability model on the likelihood of drivers reporting heckling in the past week and on the most recent trip. Panel B reports the results of an OLS regression of the likelihood of passengers reporting expressions of concern to driver/conductor on *treatment assignment status,* safety rating and the interaction of the two variables. A sample of up to 3 passengers exiting each *matatu* surveyed 8 months after recruitment is used to construct these estimates. Passengers from 22 *matatus* that could not be matched to the assignment lists are dropped leaving a total of 788 passengers. (see below on the coding of unsafe). Panel C reports the results of an ordered probit model on passenger perceptions of safety. Passengers were asked to rate the safety of the just completed trip on a scale from 1 to 10, where 1 implies no danger, 10 implies high likelihood of serious injury/death and 55 corresponds to “cannot say”. A trip is considered safe unsafe if at least one respondent reports a safety rating of 6 or higher. We recode this variable as follows: 1=Safe (a rating 1-5), 2=Cannot Say (55) and 3=Dangerous (a rating 6-10).