Working Paper Number 169
April 2009

Heckle and Chide: Results of a Randomized Road Safety Intervention in Kenya

James Habyarimana and William Jack

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 solving this collective action problem, which empowers passengers to take action. 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 at least 50 percent. Results of a driver survey eight months into the intervention suggest passenger heckling was a contributing factor to the improvement in safety.

The views expressed in this paper are those of the author and should not be attributed to the board of directors or funders of the Center for Global Development. This paper was made possible in part by support from the Australian Agency for International Development.

www.cgdev.org

James Habyarimana and William Jack. 2009. "Heckle and Chide: Results of a Randomized Road Safety Intervention in Kenya."
CGD Working Paper 169. Washington, D.C.: Center for Global Development.
http://www.cgdev.org/content/publications/detail/1421541/

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

James Habyarimana
Georgetown University

William Jack
Georgetown University

April 7, 2009

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 solving this collective action problem, which empowers passengers to take action. 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 at least 50%. 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 also 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 Nada Eissa, David Evans, Luca Flabbi, Garance Genicot, 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.

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 the health and education sectors (see Bjorkman and Svensson (2008), Svensson and Reinnika (2006) and Olken(2007)), 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] 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

[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)

(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 networks, and to enforce road laws 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 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 weekly 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: the stickers are associated with a reduction in insurance claims rates of between a half and two-thirds, from an annual baseline claims rate of about 10 percent. Further, 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 confirm that this effect is associated with consumer empowerment and action by interviewing both drivers and passengers. In particular, drivers of treated vehicles report significantly more passenger complaints than drivers of control *matatus* and conditional on experiencing a risky trip, passengers in treatment *matatus* are more likely to express concerns to their driver.

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 monitor, a speeding fine can be as much an opportunity for extortion and a source of rents as it is a Pigouvian tax. In light of the difficulty of correcting *inter*-vehicle externalities, the intervention we evaluate aims instead to correct an *intra*-vehicle externality - that between passengers and the driver - generated by features of the institutional and physical environment that induce drivers to adopt riskier behavior than passengers would likely choose. In *lieu* of the price mechanism, our stickers encourage passengers to exert social pressure on the driver, literally heckling him to take account of

[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.

[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).

the costs that his actions impose on them.

Social pressure is effective when it is social: when passengers coordinate on a particular strategy its effects might be proportionately greater. But non-cooperatively chosen actions can be inefficient from the point of view of the passengers, as when everyone sits silent hoping that someone else will chastise the driver. Our intervention is aimed at lowering the costs of action, thereby (a) increasing the likelihood that efficient choices constitute a Nash equilibrium among passengers, and (b) when multiple Pareto-comparable equilibria exist, providing a focal point that improves the chance of the more efficient one being chosen.

Micro-finance institutions have relied on social pressure to improve loan repayment rates and profitability, by making self-selected, and hence relatively homogeneous, groups liable for loans.[6] 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.[7] 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 identify 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 presents the theory on how this intervention could improve passenger action, and section 3 describes the context, data and empirical strategy. We present the results of the intervention as well as evidence for the mechanisms in section 4, and conclude in section 5.

2 Modeling 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

[6] 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.

[7] 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)).

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.

From the perspective of owners, information on actual driver behavior is virtually impossible to observe, so rewards for cautious 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 low-paid 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.[8] 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. 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.

2.1 Passenger interactions

We model passenger behavior as a non-cooperative game in which admonition of the driver is costly to the individual but has effects, in terms of driver responses, that are felt by all passengers. We propose a simple example in which two passengers, P_1 and P_2, play the following stylized game, in which each chooses a strategy of either heckling the driver or remaining quiet. The passengers face identical costs and benefits of action. This game does not examine the strategic behavior of the driver explicitly, but simply assumes some effect of passenger actions on driver behavior.

$P_1 \backslash P_2$	Heckle	Quiet
Heckle	$(\Delta - c, \Delta - c)$	$(\delta - c, \delta)$
Quiet	$(\delta, \delta - c)$	$(0, 0)$

The private cost of heckling is c. This cost can be thought of as reflecting the *ex ante* expected embarrassment associated with speaking up, or the costs of counter-heckling from the driver or other passengers who might not share a given passenger's preferences. The effect of admonition depends on how many people engage in it - the more who heckle the greater the benefit, in terms of safer driving. This benefit, which is a public good, is δ if one person heckles, and $\Delta > \delta$ if two do so. Note that as long as $\Delta > c$, which we assume throughout, the pair of strategies (H, H) Pareto dominates the pair (Q, Q), from the perspective of the passengers at least. Depending on the range of the costs and benefits of heckling, three possible games can be differentiated, with corresponding sets of equilibria, as illustrated in Panel A of Figure 1:

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

Game I: Prisoners' dilemma (high heckling cost): if $c > \Delta - \delta$, then (Q, Q) is the only equilibrium.

Game II: Coordination game (moderate heckling cost): if $c < \Delta - \delta$ but $c > \delta$, then (Q, Q) and (H, H) are both equilibria.

Game III: Prisoners' delight (low heckling cost): if $c < \Delta - \delta$ and $c < \delta$, (H, H) is the only equilibrium.

2.2 Effects of the intervention

The stickers we inserted inside the *matatus* could either increase the perceived benefits of safer driving, or reduce the costs of heckling. We admit both interpretations, in the light of apparent heterogeneity in knowledge and experience of accidents. In a survey of passengers before the stickers were designed and inserted, 11 percent of respondents reported that they or someone they knew had been in a *matatu* accident in which an injury or death occurred during the previous month. On the other hand, 55 percent knew no-one who had ever been in such a crash. Even though two of the stickers explicitly aimed to make passengers think about how bad life could be for an amputee crash survivor, we believe an equally important effect of the stickers was to empower passengers, and to legitimize the expression of their preferences. This is consistent with a reduction in the cost of heckling, c. In the model, passengers are homogeneous, drivers are identical, and we assume there is no learning by either party. We discuss the implications for our empirical results of relaxing these assumptions in section 4.

If the cost reduction associated with the stickers is big enough (so we move from Game I to Game III), then the intervention simply switches the equilibrium from (Q, Q) to (H, H) (see Panel B of Figure 1). By reducing the cost of action, the stickers could induce a move from Game I to Game II, in which case the set of equilibria is expanded from the unique inefficient equilibrium to the pair (Q, Q) and (H, H). Not only do the stickers make heckling an equilibrium, they could act as a focal point for coordinated action. Finally, the stickers might induce a switch from Game II to Game III. This switch *removes* the inefficient pair of strategies (Q, Q) from the set of equilibria, leaving the unique equilibrium (H, H). In all cases, the effect of the intervention is to increase the parameter space over which efficient heckling is observed in equilibrium.

FIGURE 1 GOES HERE

Although we do not model the strategic interaction between passengers as a group and the driver, we note that the equilbrium of that game might be characterized by no heckling even when the costs are low, if the driver knows those costs and understands that he can prevent heckling by driving safely. The actions described in the passenger interaction games above could then be thought of as expressions of a willingness to heckle – in the more complete game, we might not observe heckling on the equilibrium path, as it constitutes a credible threat. We speculate that equilbrium heckling might fall over time as the driver learns about the effects of the stickers, even while driving performance improves. In addition, the effects of message fatigue or low sticker retention could attenuate these effects. We investigate this possibility in our empirical work.

3 Context and experimental design

In this section we describe the salient features of the long distance *matatu* sector in Kenya to further motivate the model we use to rationalize the impact of our intervention. We then describe the intervention in detail and review the extent to which our experimental design was implemented in practice.

3.1 The *matatu* sector

There are perhaps 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. 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.

3.2 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"[9]; 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.[10] The chosen stickers are shown 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

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

[10] 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 what we projected to be the five most effective stickers.

front cabin.

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.[11] 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 initial recruitment experience, which revealed that vehicle lists were of variable quality, and during which non-participation rates were observed to be reasonably low, at scale-up 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 treatment/control status was maintained. Each additional observed *matatus* 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.[12] 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).[13]

The structure of the project and its objectives, were explained to each driver, as was the voluntary nature of his participation in the study.[14] Each driver in the treatment or control groups was asked to sign an informed consent form. Those selected to receive 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%. However, when we consult 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.

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

[12] 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.

[13] 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 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.

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

TABLE 1 GOES HERE

TABLE 2 GOES HERE

Actual assignments to treatment and control groups were highly correlated with the randomization. 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).

TABLE 3 GOES HERE

Imperfect adherence 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 Table 4. However, the difference in self-reported accident rates that was significant in the true assignment was narrower in the actual assignment (the rates were 1.3% and 0.6% respectively). Although we do not have a strong reason to believe that selection on the basis of those observables that show significant differences would bias our results, we check the robustness of our average treatment effect on the treated results with both intention to treat and instrumental variable estimation strategies.

TABLE 4 GOES HERE

3.3 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 table 5). Although we recognize that claims are endogenous, we do not believe they would be systematically correlated with treatment status. These data were collected for the period January 2007 through February 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. However the insurance claims data has some limitations in that we do not observe whether the vehicle involved in the claim continues to operate after the claim. Our simplifying assumption that each *matatu* continues to operate after an accident biases the result against us finding an effect of the intervention.

TABLE 5 GOES HERE

Our accident-related data were complemented by surveys of both passengers and drivers of treatment and control *matatus*, fielded in November 2008, about 8 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 examine the mechanisms by which the stickers may impact behaviors.

We are interested in estimating the 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.[15]

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*.[16]

Employing ordinary least squares, the identifying assumption for causal inference is that

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

That is, we require that unobserved factors captured by $\eta_i + \varepsilon_{it}$ are uncorrelated with the indicator for treatment. As Tables 1 and 2 demonstrate, this assumption appears to hold when TR_i corresponds to the random assignment rule. This definition of TR_i yields the well-known intent-to-treat estimator $\widehat{\beta}_3^{itt}$.

An alternative estimator is the average treatment effect on the treated, $\widehat{\beta}_3^{tot}$, in which TR_i is defined as an indicator that takes on the value of 1 if the *matatu* actually has stickers and 0 otherwise. However, in this case, the identifying assumption above is tenuous. In particular, it depends on the nature of compliance to treatment assignment. As Table 3 demonstrates, 16% of *matatus* in the control arm did not comply with their assignment. In addition, only 68% of *matatus* assigned to the treatment arm accepted all five stickers, and 16% of them accepted none. In our TOT regressions, we define $TR_i = 1$ if the vehicle accepted at least one sticker. If non-compliance is random, then $\widehat{\beta}_3^{tot}$ is a causal estimate of the effect of the intervention on the outcomes. However, if non-compliance is systematically related to unobserved factors associated with unsafe driving, then $\widehat{\beta}_3^{tot}$ is a biased estimate of the treatment effect. Table 4 presents some suggestive evidence that condition (2) might not hold, indicating that $\widehat{\beta}_3^{tot}$ could exhibit significant bias, although it is difficult to predict the direction of such.

Finally, under the assumption that compliance is endogenous to some degree, we present the results of an instrumental variables strategy in which we use the indicator for random assignment as an instrument for actual treatment status. The resulting estimator, $\widehat{\beta}_3^{iv}$, represents the local average treatment effect of the stickers on the outcome variable for the group of vehicles whose treatment status is affected by random assignment. In the results section below we present all three estimators.

[15] Anticipating depreciation and fatigue effects, a more general specification is $Y_{ik} = \alpha_0 + \sum_k \delta_k Q_{ik} + \beta TR_i + \sum_k \gamma_k Q_{ik} * TR_i + \theta X_{ik} + \eta_i + \varepsilon_{ik}$ where Q_{ik} is an indicator taking the value zero/one if quarter k is before/after the treatment of *matatu* i, and the coefficients γ_k capture the time varying effects of the intervention, post-recruitment. We present graphical evidence of a decline in the effect of the intervention where Y_{ik} is defined as the likelihood of a claims rate in quarter k.

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

4 Results

4.1 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, from 4 quarters before, to 4 quarters after recruitment.[17] Not surprisingly, quarterly claims rates are very noisy, so that before recruitment we observe moderate albeit insignificant differences across the treatment and control groups. As the figure demonstrates, while the pattern of difference in claims rates by treatment assignment status oscillates before recruitment, it has a consistent sign in the post recruitment phase. In particular, claims rates for *matatus* assigned to receive the stickers are considerably lower in the quarters after recruitment.

FIGURE 3 GOES HERE

An alternative way of presenting these results is to collapse the time before and after recruitment into two distinct time periods and to compare the change in claims rates across these two time periods for both treatment and control groups. The resulting difference-in-differences estimate captures the causal effect of the intervention on accidents if the parallel trends assumption is satisfied (another way of stating the identifying assumption in condition (2)). The results of this exercise can be presented in the form of two simple 3x3 tables, corresponding to the treatment-on-the-treated and the intent-to-treat estimators, respectively.

In the 13-15 month period before recruitment, the *matatus* in our sample experienced an average annualized claims rate of 6.47%. Over the post-recruitment period for which we have data, the average annualized claims rate among vehicles assigned to the control group increased by 3.17 percentage points, suggesting that in the absence of the intevention the average claims rate in our sample would have been 9.64%.

As Table 6 shows, even though claims rates increase significantly after the intervention for untreated vehicles, they remain constant for the treated. The difference-in-differences estimator of the effect of treatment on the treated is equivalent to a decline in the claims rate of 4.46 percentage points. If the identifying assumption of parallel trends is correct, the claims rate among treatment vehicles in the post-intervention period, had they not received stickers, would have been 10.01%, so the reduction represents a 45% drop in claims rates as a result of the intervention. This difference is significant at the 1 percent level, with a p-value of 0.0075.

TABLE 6 GOES HERE

Because patterns of non-compliance to assignment could be systematically related to accident rates, we repeat this exercise in Table 7, this time defining TR_i to indicate *random assignment* to treatment. The ITT estimate is virtually unchanged from the TOT estimate, being 4.47 percentage points, or a 44% fall from a projected rate among those assigned to treatment of 10.16%, and still highly significant (p-value 0.0085). A positive difference between $\widehat{\beta}_3^{tot}$ and $\widehat{\beta}_3^{itt}$ would suggest a pattern of non-compliance in which the more risky drivers assigned to treatment opted not to accept the stickers and the less risky drivers assigned to control accepted the stickers. However, we are unable to reject the null hypothesis that the point estimates are the same.

[17] The horizontal axis in this figure measures time since recruitment, not calendar time. As recruitment took eight to ten weeks, calendar time varies by vehicle. As of the time of writing, we were not able to obtain data on all recruited vehicles for the fourth post-recruitment quarter. As mentioned earlier, we assume that vehicles continue to operate after an accident so that the denominator remains the same before and after recruitment.

TABLE 7 GOES HERE

Table 8 restates these TOT and ITT double difference estimates (columns 1 and 3 respectively), but in addition presents estimates of the same effects when SACCO fixed effects are included (columns 2 and 4, respectively). The point estimates are virtually unchanged.

TABLE 8 GOES HERE

In columns (5) and (6) of Table 8 we present the instrumental variable estimates in which we instrument for actual treatment status using the random assignment. The local average treatment effect of 6.5 percentage points is nearly 50% larger than the ITT estimator. Relative to the projected claims rate, the LATE estimator suggests a decline in the rate of accidents of as much as 65% associated with the treatment. While the usefulness of some IV results is legitimately questioned in the face of weak instrument problems and heterogeneity, we believe our strong first stage and high compliance rates make this a credible estimate of the impact of the stickers. Nearly two-thirds of the accidents that would otherwise have occurred are avoided.

In 278 of the 319 claims events in our data (about 87 percent) that we could classify,[18] the *matatu* driver is recorded as being at fault. Using these data, Table 9 presents an ITT double difference estimate of the impact of our intervention on driver-at-fault claims. The point estimate of -4.10 percentage points remains highly significant (p-value 0.0070) and represents a 46% reduction in driver-at-fault claims below the projected base.

TABLE 9 GOES HERE

Finally, the intervention we evaluate appears to reduce serious accidents. Our data include 206 claims with at least one injury or death. Using this as an outcome variable, we repeat our ITT difference in differences analysis in Table 10. Again, the point estimate of 3.35 percentage points is highly significant (p-value = 0.0079) and large, representing a 50% reduction in such accidents from the projected base of 6.65 percent.

TABLE 10 GOES HERE

4.2 Sustainability

The effectiveness of the stickers in solving the collective action problem we identify could vary over time. The most obvious reason is depreciation of the stickers, which might be physically removed, or simply fade and deteriorate with extended exposure to dusty country roads and repeated washing. But the stickers could have some longer-term effect on individuals who see them. The majority of the *matatus* in our sample are operated by the same driver over time, so we might expect drivers to exhibit some learning, habit formation, or other behavioral effects of a long-run nature. On the other hand, the exposure of passengers to the treatment is less uniform; at later dates after the intervention, some riders will be seeing the stickers for the first time, while others will have been exposed potentially many times, depending on the frequency of their trips, and their use of treated and untreated vehicles. Observed behavioral change among passengers might then be somewhat slower.

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

To illustrate the temporal effects of the treatment, we present quarterly estimates of the differential claim probability between *matatus* assigned to the treatment and control groups. Due to the low frequency of events in each group, the standard errors we calculate for quarterly data are relatively large; nonetheless we believe this exercise provides useful information about the sustainability of the intervention.

Figure 4 shows the differential probability of a claim being filed in each of the four quarters prior to the intervention, and the four quarters afterwards. While at this level of disaggregation none of the quarterly differentials are statistically significant, there is a clear, steady, decline in the magnitude of quarterly point estimates post-recruitment. *Matatus* assigned to the treatment are about 1.25 percentage points less likely to file a claim during the first post-recruitment quarter (i.e., about 5 points on an annualized basis). This falls to about 0.9 points in the second quarter, and 0.5 points in the third.

FIGURE 4 GOES HERE

This figure offers suggestive evidence of a reduction in the effectiveness of the intervention. However, based on our survey results eight months after recruitment, the reduction in the impact of the stickers appears to match the reduction in the number of vehicles with stickers: that is, our evidence is consistent with a situation in which the messages maintain their salience, conditional on remaining in view of the passengers. As Table 12 shows, the share of *matatus* with all five stickers fell to about 40% of the number at recruitment after eight months, which is consistent with the reduction in the quarterly claim differential over the same period, from 1.25 to 0.5 percentage points, although our estimates lack precision.

FIGURE 5 GOES HERE

Figure 5 presents a similar pattern of declining effectiveness post-recruitment, as measured by the difference in probability of a claim being filed in which the driver was at fault. Compared to Figure 4, the decline is less steep over the three quarters for which we have complete data. Finally, Figure 6 reports similar estimates for claims that involved either an injury or a fatality. The effects are smaller by this measure, and even less precisely estimated, although it appears the impact of the stickers had all but disappeared by the third quarter post-recruitment.

FIGURE 6 GOES HERE

4.3 Heterogeneous Treatment Effects

The mechanism by which our treatment might affect driver behavior is potentially complex, as it involves a number of decisions-makers. The simple theory described in Section 2 assumed a homogenous group of passengers, and a positive and uniform response to heckling across all drivers. However, passengers and drivers can each differ in a number of relevant ways, suggesting plausible heterogeneous treatment effects. This possibility arises not only because individual passengers and drivers might exhibit heterogeneous underlying characteristics, but because the characteristics of other agents are difficult to observe. For example, a passenger might be illiterate, so might be motivated little by the stickers; and even if he could read, he might not know if the other passengers can read, and if his objections will be backed up. Similarly, the reaction of the driver to heckling might depend on his personality, and the impact on the safety of the trip could be correlated with this. Thus, “good” drivers might slow down in response to heckling, while “bad” drivers might respond, for example, by overtaking more aggressively.

Finally, due to the strategic nature of the interactions among passengers, and between them and the driver, the beliefs that individuals hold about the characteristics of other players are important. If a passenger's experience suggests to him that drivers in general are very likely to respond negatively to heckling, or that other passengers are unlikely to join him, then irrespective of the costs of heckling, he will be cautious. If, on the other hand, the passenger has little experience of the stickers, he might be willing to experiment in order to learn about both the responsiveness of other passengers, and the reaction of the driver. Heterogeneity of underlying characteristics, as well as beliefs about those characteristics thus yield, at least in theory, a myriad of potential effects of the stickers on safety.

These observations have implications for the interpretation of our empirical results. In particular, a positive intention-to-treat estimator does not imply a positive treatment effect across the distribution of passenger and driver types. Furthermore, to the extent that compliance to random assignment holds for a subset of driver types, the instrumental variable estimate may correspond to the treatment effect of a small and in policy terms unimportant group of drivers. However, given the size of our ITT estimates, and the high compliance rates achieved, these problems of interpretation do not appear to place significant limitations on our analysis.

4.4 Robustness of Treatment effect

In this sub-section we present two robustness tests for the main results outlined above. In general, difference-in-differences estimates rely on the strong assumption of parrallel trends in the outcomes in the absence of the intervention. If this assumption does not hold, the measured treatment effect reflects trend differences between the treatment and control groups. Our randomization should have eliminated such differences, but nevertheless we perform a falsification test in which we create a synthetic recruitment date for every *matatu* that is exactly one year before the actual recruitment date. Using insurance claims data for the two year window around this synthetic recruitment date, we carry out a difference-in-differences estimation strategy to examine whether there are trend differences between the treatment and control vehicles.[19]

TABLE 11 GOES HERE

The results of this exercise are shown in Table 11. The difference-in-differences estimate for this placebo test is positive and insignificant. Additional evidence using a quarterly probit that examines differences in the likelihood of quarterly claims is shown in Figure 7. This figure exhibits no systematic differences between treatment and control *matatus* either before or after the synthetic recruitment date. Both of these results suggest that the main results presented above are unlikely the consequence of trend differences in the two groups.

FIGURE 7 GOES HERE

4.5 Evidence on mechanisms of change

The theory presented in section 2 suggests that passengers traveling in *matatus* with stickers will be more likely to voice their concerns over bad driving. To investigate if this could be leading to the observed differential in claims rates identified above, we analyze data from a survey of drivers, plus up to three of

[19] In running this falsification test we make the simplifying assumption that all matatus recruited between March-May 2008 were in operation throughout 2006 and 2007.

their passengers, in 284 vehicles fielded in November 2008, about 8 months after recruitment.[20] We face two difficulties in detecting evidence for the mechanisms underlying our results. First, even if the stickers are effective, 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 many of the treatment vehicles had lost some or all of their stickers. Table 12 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 12 GOES HERE

Table 13 reports evidence of heckling from the survey of drivers and reports of passenger safety ratings. The first panel presents OLS coefficients on two indicators, the first indicating whether the vehicle had received at least one sticker at recruitment, and the second that it had at least one sticker at the time of the survey.[21] The second panel presents results from regressions in which we instrument for both accepting stickers at recruitment and having them at the time of the survey. We use random assignment status at recruitment as an instrument for initial treatment, and the gender of the recruiting enumerator as an instrument for having retained at least some stickers through November 2008. Our instrument for sticker retention is disappoiningly weak, and our IV estimates are large and insignificant. However the OLS results do provide some evidence of heckling. The effect of stickers on driver-reported accidents since recruitment, reported in column (1), is in the right direction, albeit imprecisely estimated. However, columns (2) and (3) show that compared to vehicles with no stickers at recruitment, drivers of vehicles with stickers in November were 10.4 percentage points more likely to have experienced passenger heckling during the past week, and 6.1 percentage points more likely to have experienced it during the most recent trip. Joint tests of significance for these two indicators suggest significant explanatory power for passenger helcking in the past week and borderline significance (p-value 0.11) for the most recent trip. These differences are compared with low heckling rates in *matatus* without stickers, 5.6% and 3.8% respectively. Drivers of vehicles with stickers are thus about three times more likely to report heckling.

TABLE 13 GOES HERE

Our results on passenger rating of the safety of the most recent trip in column (4) do not provide evidence of drivers anticipating heckling and driving more safely. 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. Nearly 45% of the respondents who reported that they "could not say" were dropped from the analysis. We define a trip to be reported as safe if the passenger reports a rating less than 4. About two thirds of all passengers in the control *matatus* rated the most recent trip as safe according to this definition. As the OLS estimates suggest, passengers in *matatus* with stickers are nearly 4 percentage points less likely to report a safe trip. While the sign of this estimate suggests that stickers might make *matatu* passengers feel less safe (a salience effect consistent with one of the proposed mechanisms), it is imprecisely estimated.

Passenger reports of heckling have the potential to provide further evidence on the mechanisms that might lead to our results. Sampled passengers were asked to report if they or any other passengers had said

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

[21] The full impact of the stickers is thus the sum of the two coefficients.

something to the driver/conductor about reckless driving behavior. We divide the data into two categories depending on whether at least one passenger had rated the safety of the trip as dangerous (a rating of 6 or higher). In addition to corroborating the mechanisms outlined, this dichotomy allows us to investigate whether the stickers generate inefficient levels of heckling when there are no risks of accidents.

Table 14 presents the results of this exercise. Panel A presents the results for heckling by any passenger, panel B presents results for heckling by the respondent, and panel C investigates the likelihood of social pressure, that is, heckling by multiple respondents. Three estimates of the proportion of passengers reporting heckling are presented for passengers in vehicles with no stickers at recruitment, in vehicles that received, but no longer have stickers, and in those that had retained their stickers up until the time of the survey. Among trips considered safe, passenger reports of heckling are common, with 50 percent of respondents reporting that a passenger had heckled the driver on the most recent trip in vehicles that received no stickers. Heckling rates are very similar among vehicles that had stickers (44%) and those currently with stickers (47%). Assuming that measurement error is not correlated with stickers (a strong assumption given the content of the intervention), we do not find evidence of excessive heckling in treatment vehicles that had just completed a safe trip. Among those trips considered risky by at least one passenger we find evidence of a 50% higher rate of heckling among passengers in *matatus* with stickers (54% vs 36% in the control).

TABLE 14 GOES HERE

Turning to panel B of the table, which reports the rates of heckling by the survey respondent him/herself, we find similar evidence for the lack of excessive heckling when trips are safe, and differential heckling when trips are unsafe. In particular we find that passengers in *matatus* with stickers are nearly 3 times as likely to heckle the driver as passengers in *matatus* with no stickers at recruitment. We note that the rate of heckling among passengers in *matatus* that had stickers at recruitment, but no longer do so, is inconsistent with a no-learning effect of the stickers.

Finally we investigate the extent to which the stickers provide a focal point for more than one passenger to heckle the driver. We define our outcome as the likelihood that two or more respondents heckled the driver during the just completed trip. We do not know if such multi-person heckling occurred in response to the same dangerous event, or if each heckler responded to a different incident, so we cannot definitively say if the reports correspond precisely to the kind of coordinated social pressure outlined in Section 2, although we believe our results support this interpretation. The sample is divided into safe and unsafe trips, as above, and the results are presented in panel C of Table 14. Again we find no differences in multiple reports of concern among trips that are considered safe. However, we find large differences among trips considered risky. A vehicle with stickers is nearly three times as likely as a vehicle with no stickers at recruitment to have multiple responses of concern about the driver's behavior.

5 Conclusions

We have presented evidence that a very cheap intervention can overcome a potentially catastrophic collective action problem 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 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- and passenger-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.

Solving the collective action problem among passengers 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 a passenger-based intervention. 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 duration over which such interventions are effective, and how frequently messages need to be updated.

6 References

References

[1] Armendariz de Aghion, Beatriz and Jonathan Morduch (2000): "Microfinance beyond group lending," *The Economics of Transition*, 8 (2): 401 - 420.

[2] 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.

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

[4] Bardhan, Pranab, (2000), "Irrigation and Cooperation: An Empirical Analysis of 48 Irrigation Communities in South India," *Economic Development and Cultural Change* 48, pp. 847-65

[5] 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

[6] 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.

[7] 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.

[8] 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*

[9] Caballero, María Cristina (2004): "Academic turns city into a social experiment," *Harvard University Gazette*, march 11, 2004. http://www.hno.harvard.edu/gazette/2004/03.11/01-mockus.html

[10] Camerer, Colin F., and Dan Lovallo. March 1999. "Overconfidence and Excess Entry: An Experimental Approach." *American Economic Review*, 89, 306-18

[11] 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.

[12] Diamond, Peter A., 1971. "A model of price adjustment," *Journal of Economic Theory*, vol. 3(2):156-168.

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

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

[15] 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.

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

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

[18] Khwaja, Asim 20048, "Can Good Projects Succeed in Bad Communities? Collective Action in the Himalayas," unpublished manuscript, Harvard University

[19] 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.

[20] Lovallo D., & Kahneman D. (2003). "Delusions of success: How optimism undermines executives' decisions". *Harvard Business Review*, 81 , 56-63

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

[22] 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.

[23] 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

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

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

[26] Morduch, Jonathan (1998): "Does microfinance really help the poor? New evidence from flagship programs in Bangladesh," working paper, Wagner Graduate School of Public Service, NYU, http://www.nyu.edu/projects/morduch/documents/microfinance/Does_Microfinance_Really_Help.pdf

[27] 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.

[28] Okten, Cagla and Una Okonkwo Osili, 2004, "Contributions in Heterogeneous Communities: Evidence from Indonesia," *Journal of Population Economics* 17 (December), pp. 603-626

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

[30] Pitt, Mark (1999): "Reply to Jonathan Morduch's Does microfinance really help the poor? New evidence from flagship programs in Bangladesh," working paper, Brown University, http://www.pstc.brown.edu/~mp/reply.pdf

[31] 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.

[32] Svensson, J. and Reinnika, R. 2006. "The power of information: Evidence from a newspaper campaign to reduce capture" *Quarterly Journal of Economics*

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

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

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

Appendix: Stickers inserted in treatment *matatus*

Hey, if he's driving recklessly, will you arrive?
BE AWAKE. BE STEADY. SPEAK UP!

Hey, will you complain after he causes an accident?
BE AWAKE. BE STEADY. SPEAK UP!

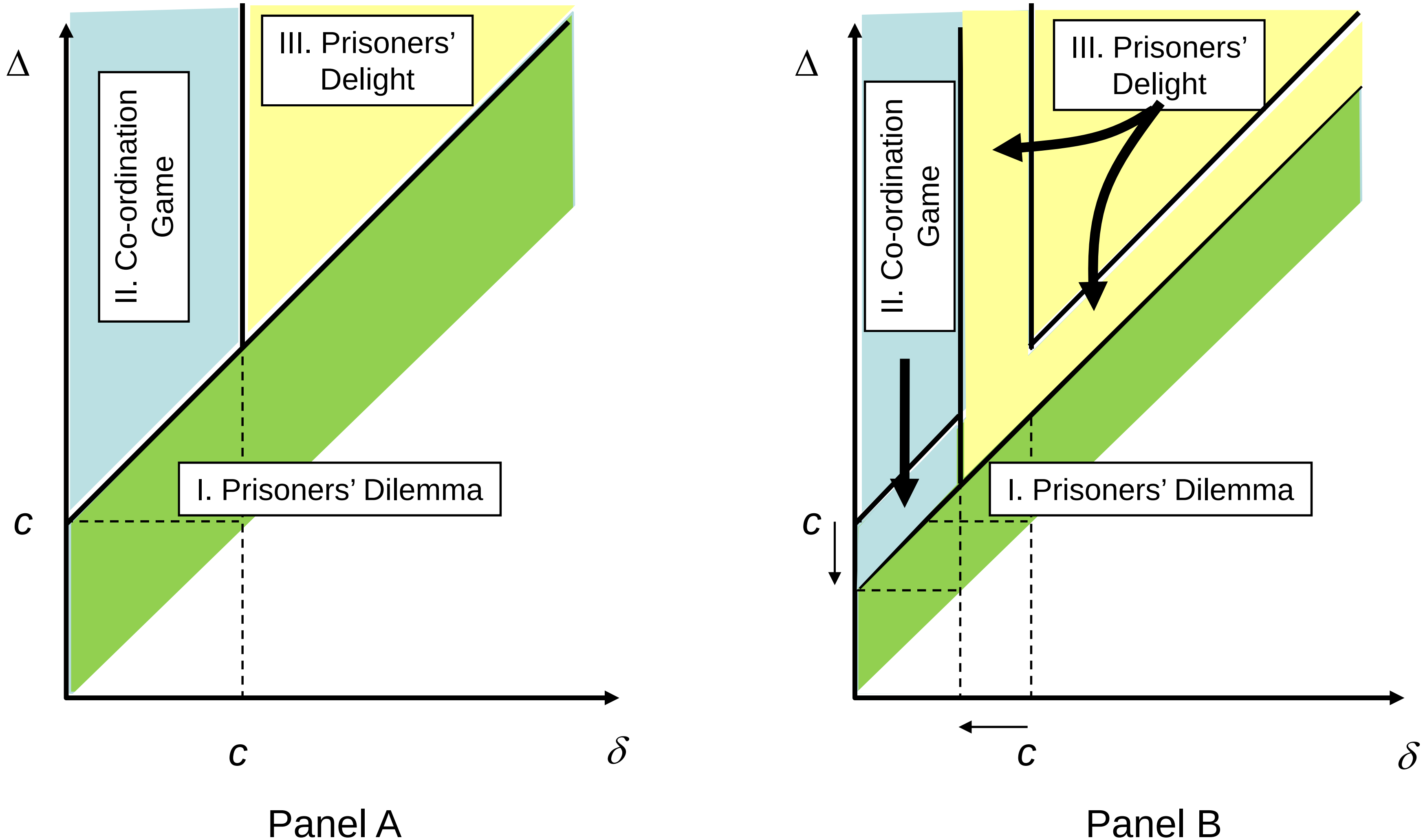
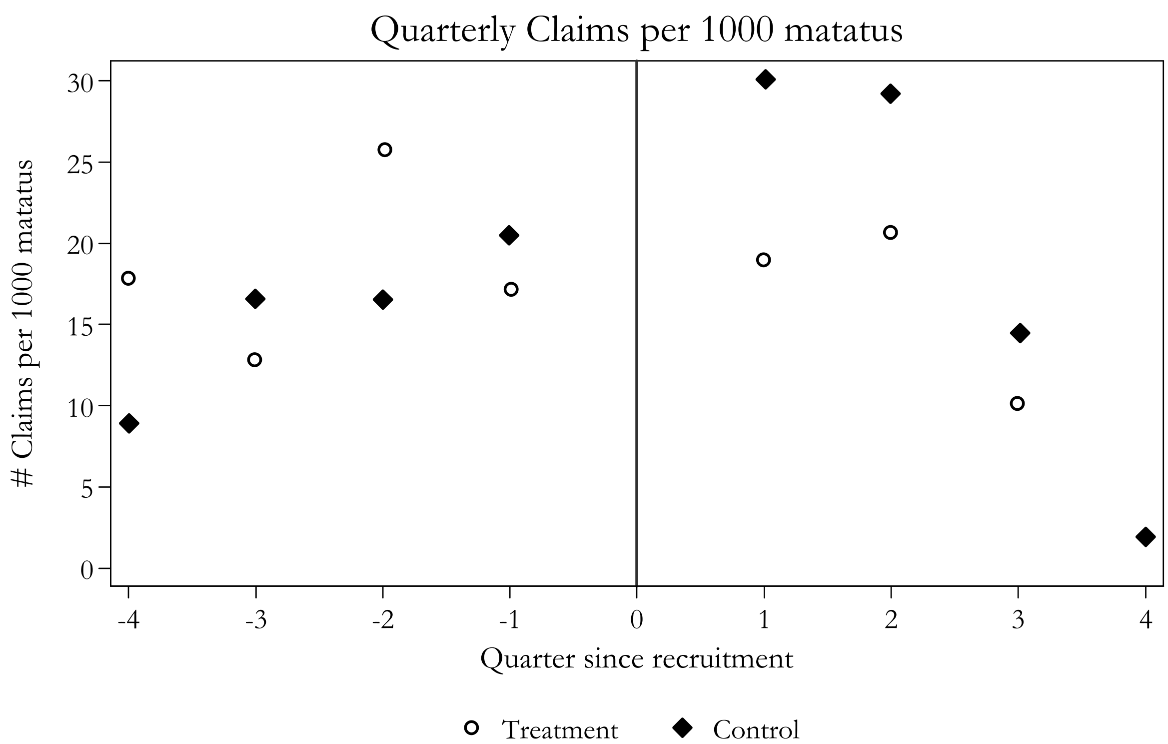


Figure 1: Panel A: Costs and benefits of heckling and resulting equilibria. Panel B: The impact of a fall in the cost of heckling on equilibria

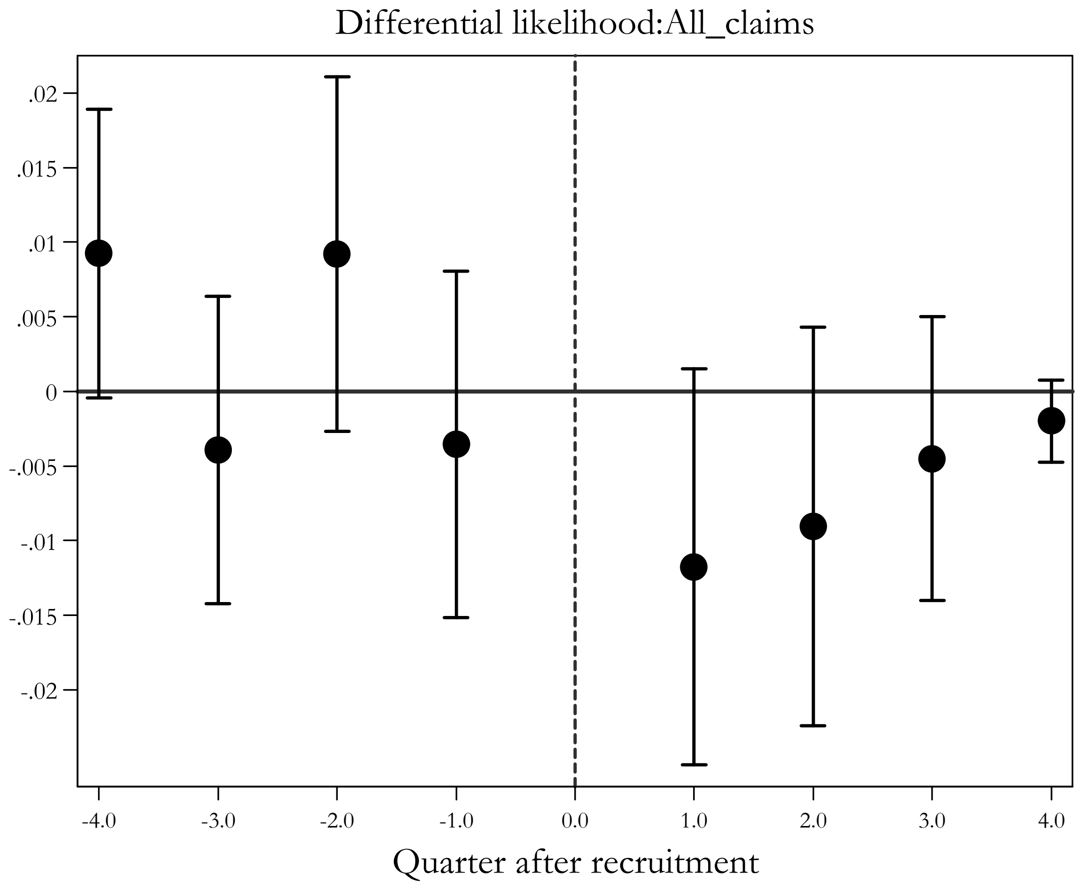
Figure 2: Major towns served by sampled long-distance *matatus.*

Figure 3: Number of events per 1,000 *matatus* and quarter before/after recruitment



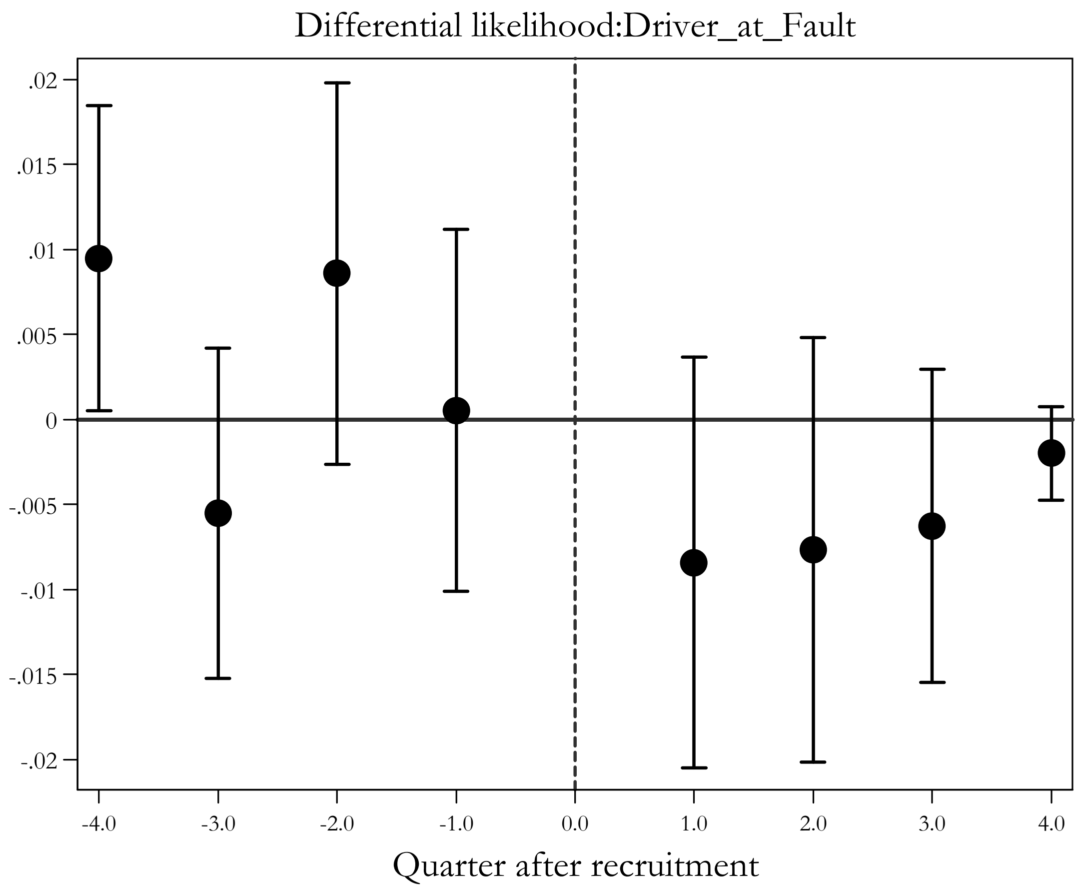
Note: The figure presents the number of insurance claims-generating incidents per 1,000 *matatus* for each quarter since recruitment, by treatment assignment. The vertical line at quarter 0 defines the *matatu*-specific date of recruitment which varies between March-April of 2008. All insurance claims matched to the experimental sample from January 2007 to February 2009 are used to construct this figure. We make a simplifying assumption that *matatus* continue to operate after a claim event and were operating throughout this period.

Figure 4: Differential Likelihood of Claim by quarter since recruitment – All Claims



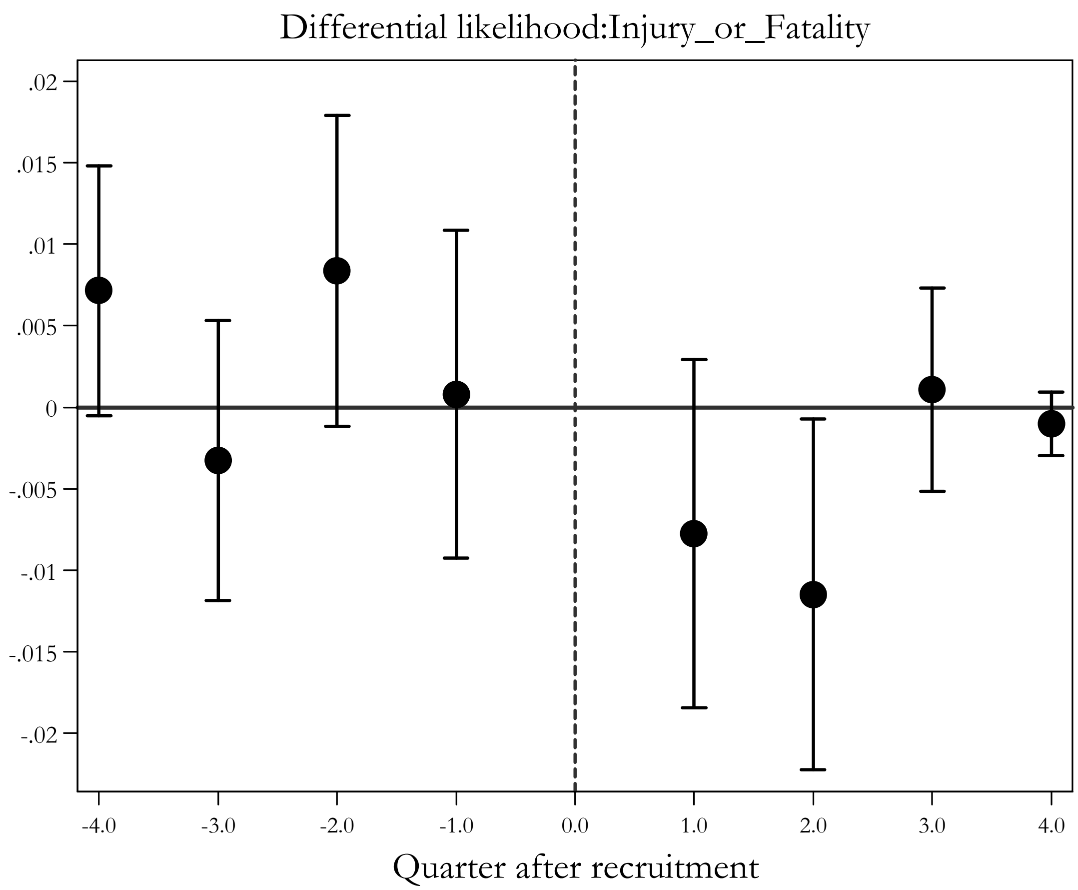
Note: The figure presents the results of marginal probit estimation of the relative likelihood of an insurance claim generating event by quarter since recruitment. The smile plot represents the differential likelihood of an event (with 95% confidence interval) for *matatus* assigned to the treatment group relative to those assigned to the control group for every quarter. The dashed line at quarter 0 defines the *matatu*-specific date of recruitment which varies between March-April of 2008. *All* insurance claims between January 2007 and February 2009 are used to construct this figure. We make a simplifying assumption that *matatus* continue to operate after a claim event and were operating throughout this period.

Figure 5: Differential Likelihood of Claim by quarter since recruitment – Driver at Fault Claims Only



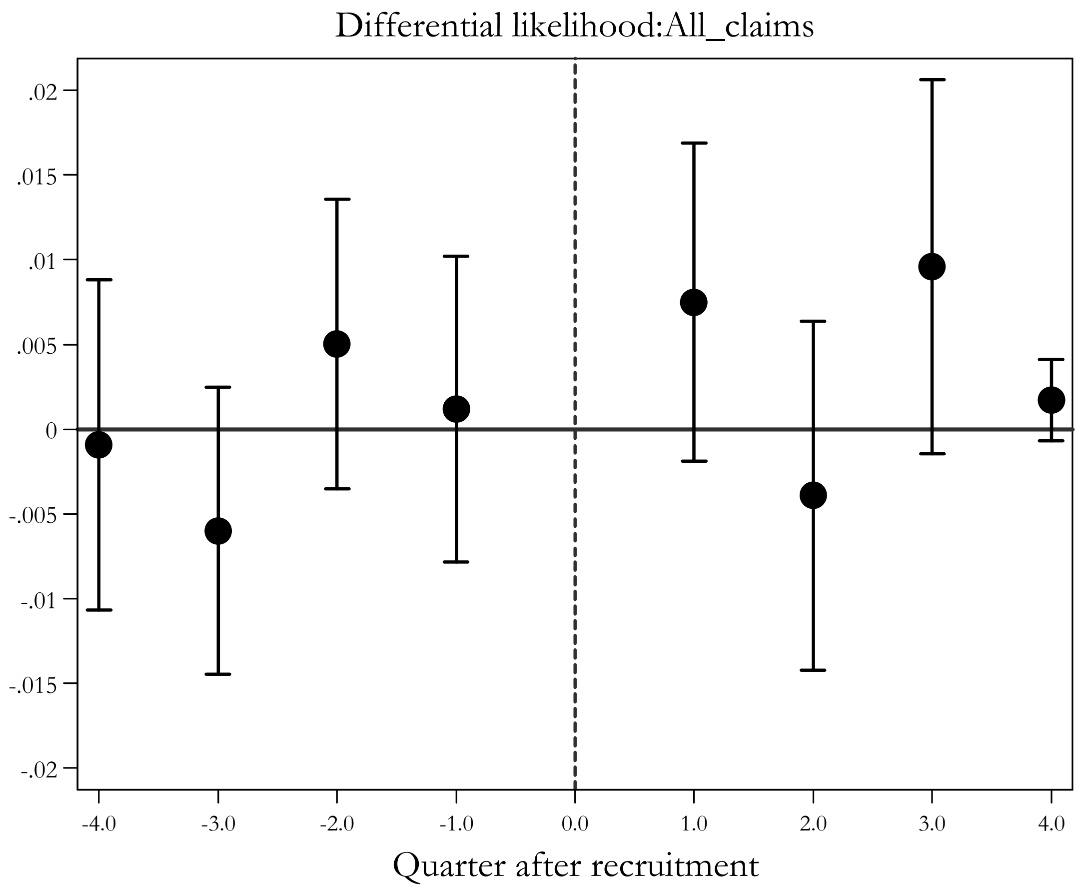
Note: The figure presents the results of marginal probit estimation of the likelihood of a driver-at-fault insurance claim generating event by quarter since recruitment. The smile plot represents the differential likelihood of an event (with 95% confidence interval) for for *matatus* assigned to the treatment group relative to those assigned to the control group for every quarter. The dashed line at quarter 0 defines the *matatu*-specific date of recruitment which varies between March-April of 2008. *Only* insurance claims between January 2007 and February 2009 for *which the driver is deemed to have been at fault* are used to construct this figure. We make a simplifying assumption that *matatus* continue to operate after a claim event and were operating throughout this period.

Figure 6: Differential Likelihood of Claim by quarter since recruitment – Injury or Fatality Claims



Note: The figure presents the results of marginal probit estimation of the likelihood of an event generating an insurance claim involving an injury or fatality by quarter since recruitment. The smile plot represents the differential likelihood of an event (with 95% confidence interval) for *matatus* assigned to the treatment group relative to those assigned to the control group for every quarter. The dashed line at quarter 0 defines the *matatu*-specific date of recruitment which varies between March-April of 2008. *Only* insurance claims between January 2007 and February 2009 involving *an injury or fatality* are used to construct this figure. We make a simplifying assumption that *matatus* continue to operate after a claim event and were operating throughout this period.

Figure 7: Falsification Test: Differential likelihood of claim by quarter since *placebo* recruitment



Note: The figure presents the results of marginal probit estimation of the likelihood of an insurance claim generating event by quarter since a *placebo* recruitment date that corresponds to exactly one year since actual recruitment. The smile plot represents the differential likelihood of an event (with 95% confidence interval) for *matatus* assigned to the treatment group relative to those assigned to the control group for every quarter. The dashed line at quarter 0 defines the *placebo matatu*-specific date of recruitment which varies between *March-April 2007*. *All* insurance claims between *January 2006 and December 2007* are used to construct this figure. We make a simplifying assumption that *matatus* continue to operate after a claim event and were operating throughout this period.

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
	(0.00)	(0.00)		(0.00)
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)		(0.00)
Insurance claim filed in last 12 months before recruitment	0.059	0.070	No	0.055
	(.007)	(.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 permanent	0.72	0.70	No	0.71
drivers	(0.04)	(0.04)		(0.03)
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
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
	(0.00)	(0.00)	
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	0.060	0.069	No
	(.007)	(.007)	
Number of observations	1035	1126	

Notes: Standard errors in ( ); Medians in [ ]. The table presents mean/median of 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.

Table 5: Market Share for Third Party Insurance

Company Name	Percent
Amaco	4.3
Blue Shield	40.9
Direct Line	38.9
Standard Assurance	7.1
Other	8.8
Total	100.00

Notes: The table presents the share of *matatus* in a random sample of our study *matatus* covered by the four companies that provide the insurance claims data used as the primary outcome. The sample is used for this table is based on a random sample of 284 *matatu* drivers who were surveyed about 6 months after recruitment.

Table 6: Difference-in-Differences - By Actual Treatment

Actual treatment status	Before	After	Difference (After –Before)
Control (No stickers)	0.0601 (.0073)	0.0913 (.0106)	0.0312 (.0129)
Treatment (Stickers)	0.0689 (.0071)	0.0554 (.0079)	-.0135 (.0106)
Difference (Treatment-Control)	0.0087 (.0102)	-0.0359 (.0132)	**-0.0446** **(.0167)**

Notes: Standard errors in ( ); The table presents annualized average claims rates by *actual treatment status* before and after recruitment. All insurance claims matched to the experimental sample from January 2007 to February 2009 are used to construct the claims rates. We make a simplifying assumption that *matatus* continue to operate after a claim event and were operating throughout this period. 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 7: Difference-in-Differences - By Assignment

Assignment	Before	After	Difference (After-Before)
Control	0.0587 (.0075)	0.0905 (.0111)	0.0317 (.0134)
Treatment	0.0699 (.0069)	0.0571 (.0076)	-0.0128 (.0103)
Difference (Treatment –Control)	0.0113 (.0102)	-0.0334 (.0135)	**-0.0447** **(.0168)**

Notes: Standard errors in ( ); The table presents annualized average claims rates by *treatment assignment* before and after recruitment. All insurance claims matched to the experimental sample from January 2007 to February 2009 are used to construct the claims rates. We make a simplifying assumption that *matatus* continue to operate after a claim event and were operating throughout this period. 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 8: Regression Results

	OLS: Actual Treatment on Treated		Reduced Form: Intent to Treat		Instrumental Variables	
	(1)	(2)	(3)	(4)	(5)	(6)
Post	0.0312	0.0320	0.0318	0.0326	0.0420	0.0430
	(0.0154)*	(0.0124)*	(0.0173)*	(0.0141)*	(0.0113)*	(0.0091)**
Treatment	0.0088	0.0107	0.0113	0.0103	0.0165	0.0146
	(0.3873)	(0.3058)	(0.2673)	(0.3148)	(0.2674)	(0.3383)
Post*Treatment	-0.0446	-0.0454	-0.0447	-0.0455	-0.0654	-0.0667
	(0.0075)**	(0.0062)**	(0.0080)**	(0.0066)**	(0.0080)**	(0.0067)**
Constant	0.0601	0.0393	0.0587	0.0394	0.0561	0.0371
	(0.0000)**	(0.0034)**	(0.0000)**	(0.0032)**	(0.0000)**	(0.0117)*
Controls for SACCO		X		X		X
Observations	4322	4318	4322	4318	4322	4318
R-squared	0.0025	0.0167	0.0023	0.0167	0.0021	0.0162
First Stage F-stat					2421.33	2364.44

Notes: Robust **p-values** 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 9: Difference-in-Differences: Driver at Fault - By Assignment

Assignment	Before	After	Difference (After-Before)
Control	0.0494	0.0770	0.0276
	(.0067)	(.0098)	(.0119)
Treatment	0.0624	0.0490	-0.0134
	(.0066)	(.0069)	(.0095)
Difference (Treatment –Control)	0.013	-0.028	**-0.041**
	(.0094)	(.0120)	**(.0152)**

Notes: Standard errors in ( ); The table presents annualized average claims rates by *treatment assignment* before and after recruitment. Only insurance claims matched to the experimental sample from January 2007 to February 2009 for which the driver is deemed to have been at fault are used to construct the claims rates. We make a simplifying assumption that *matatus* continue to operate after a claim event and were operating throughout this period. 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 10: Difference-in-Differences: Claims With Injury or Fatality - By Assignment

Assignment	Before	After	Difference (After-Before)
Control	0.0371 (.0056)	0.0546 (.0079)	0.0175 (.0097)
Treatment	0.0490 (.0058)	0.0330 (.0057)	-0.016 (.0081)
Difference (Treatment –Control)	0.0119 (.0081)	-0.0216 (.0097)	**-0.0335** **(.0126)**

Notes: Standard errors in ( ); The table presents annualized average claims rates by *treatment assignment* before and after recruitment. Only insurance claims matched to the experimental sample from January 2007 to February 2009 with an injury or fatality are used to construct the claims rates. We make a simplifying assumption that *matatus* continue to operate after a claim event and were operating throughout the pre- and post-recruitment period. 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 11: Falsification Test: Difference-in-Differences - By Assignment

Assignment	Before	After	Difference (After-Before)
Control	0.0427 (.0061)	0.0542 (.0091)	0.0115 (.011)
Treatment	0.0464 (.0057)	0.0745 (.0095)	0.0281 (.0111)
Difference (Treatment –Control)	0.0037 (.0083)	0.0203 (.0132)	**0.0166** **(.0156)**

Notes: Standard errors in ( ); The table presents annualized average claims rates by *treatment assignment* before and after a *placebo* recruitment date that corresponds to exactly one year before actual recruitment. All insurance claims matched to the experimental sample from January 2006 to December 2007 are used to construct the claims rates. We make a simplifying assumption that *matatus* continue to operate after a claim event and were operating throughout the *synthetic* pre- and post-recruitment period. 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 12: 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 13: Testing for Mechanisms: Driver responses and safety rating of most recent trip

	(1)	(2)	(3)	(4)
	Driver reports accident since recruitment	Driver reports heckling (Past week)	Driver reports heckling (Most recent trip)	Safety rating (Indicator for very little danger during trip; safety rating < 4)
	Panel A: Ordinary Least Squares			
Treated At Recruitment	-0.019	-0.014	-0.038	0.065
	(0.013)	(0.033)	(0.019)*	(0.053)
Remained treated in November 2008	0.012	0.118	0.099	-0.105
	(0.012)	(0.048)*	(0.033)**	(0.059)+
	Panel B: Instrumental Variables			
Treated At Recruitment	-0.224	-0.404	-0.562	-0.496
	(0.476)	(0.983)	(1.109)	(0.721)
Remained treated in November 2008	0.393	0.837	1.030	0.891
	(0.911)	(1.854)	(2.101)	(1.392)
Observations	259	259	258	418
R-squared (OLS)	0.01	0.03	0.03	0.01
F-stat stickers matter(OLS)	0.12	4.99	2.55	0.54
p-value stickers matter(OLS)	0.73	0.03	0.11	0.46
Mean of dependent variable: Controls only	0.019	0.056	0.038	0.644

Notes: Robust standard errors in parentheses. + significant at 10%; * significant at 5%; ** significant at 1%. Table reports the estimates of linear probability models of post treatment outcomes. Ordinary least squares are reported in the first panel while IV estimates are reported in panel B. Assignment to treatment is used as the instrument for treatment at recruitment while an indicator for female recruiter is used as an instrument for treated in November. The mean of the dependent variable shown is calculated for control *matatus* only. Sample restricted to *matatus* surveyed 8 months after recruitment.

Table 14: Testing for Mechanisms: Likelihood of self-reported passenger heckling

	No Stickers at Recruitment	Stickers Recruitment
Panel A: Proportion of respondents reporting *if any passenger* expressed concern		
	None of the respondent report dangerous trip	
Sticker Retention	0.50	
No stickers in November		0.44
Stickers in November		0.47
	At least one respondent reports dangerous trip	
Sticker Retention	0.36	
No stickers in November		0.38
Stickers in November		0.54
Panel B: Proportion of *respondents* expressed concern		
	None of the respondents report dangerous trip	
Sticker Retention	0.23	
No stickers in November		0.22
Stickers in November		0.17
	At least one respondent reports dangerous trip	
Sticker Retention	0.12	
No stickers in November		0.29
Stickers in November		0.33
Panel C: Proportion of vehicles with multiple respondent expressing concern		
	None of the respondents report dangerous trip	
Sticker Retention	0.13	
No stickers in November		0.09
Stickers in November		0.08
	At least one respondent reports dangerous trip	
Sticker Retention	0.08	
No stickers in November		0.29
Stickers in November		0.25

Notes: Table reports the mean proportion of passengers reporting expressions of concern to driver/conductor by *treatment status at the time of the survey*. 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 785 passengers. Passengers were asked to rate the safety of the just completed trip on a scale from 1 to 10, where 1 implies no danger, and 10 implies high likelihood of serious injury/death. A trip is considered dangerous if at least one responded reports a safety rating of 6 or higher. About 10% of *matatu* trips were rated as dangerous by this definition.