Final report ETH-25107

February 2026

The effect of incentive structures on driver behavior and urban road safety in Addis Ababa

Tsega Adego
Eyoual Demeke
Nachiket Shah
Afras Sial

IGC

DIRECTED BY LSE UNIVERSITY OF OXFORD

FUNDED BY UKaid from the British people

Final Report ETH-25107

The Effect of Incentive Structures on Driver Behavior and Urban Road Safety in Addis Ababa

Tsega Adego, Eyoual Demeke, Nachiket Shah, Afras Sial

Policy Studies Institute and University of California, Berkeley

February 2026

Table of Contents

1. Executive Summary

1.1 Purpose

Poor urban road safety is a pervasive issue across Sub-Saharan Africa, and one that will only continue to grow as urbanization accelerates across the region. A key source of poor road safety in African cities is the prevalence of informal, privately-owned minibuses, which account for the majority of public transportation. This pilot study designs and implements an RCT to test the effectiveness of financial and non-financial incentive programs in improving the reckless driving behaviour of minibus drivers in Addis Ababa, Ethiopia.

1.2 Key findings

- There is some suggestive evidence that the non-financial incentive reduced the number of speeding and sudden braking violations incurred by drivers, although the magnitude of the effect is small
- There was no real impact of the financial incentive on driver road safety outcomes
- Drivers are more willing to pay to be a part of the financial incentive program, than the non-financial program. These patterns do not change much over time, as drivers are exposed to the programs.
- Drivers are generally overconfident in their beliefs of how safe they are at the beginning of the study. By the end of the study, drivers in both groups are more confident about how safely they believe they drive.

1.3 Key conclusions

As this is a pilot study of only 50 drivers, all results must be treated as very preliminary, as the standard errors on all estimates/findings are quite large. Nevertheless, the broad patterns found in this pilot indicate that running a scaled-up version of these programs on a larger sample of drivers could be promising, as these incentives are feasible, understood by drivers, and seem to have some impact on their road safety behaviour.

2. Introduction

Road traffic accidents (RTAs) are a global health issue with a reported yearly death rate of 1.19 million persons, and 20 to 50 million injuries and disabilities. Despite less than 60% of vehicle ownership, 92% of these deaths occur in low and middle-income countries, with Africa having the world's highest estimated road traffic fatalities of over 300,000 people/year (WHO, 2026).

In Ethiopia, RTAs were the second most prevalent type of accident, accounting for around 43% of fatalities, with an accident rate of 163 per 100,000 people per year (Abegaz & Gebremedhin, 2019). In Addis Ababa alone, there were 408 reported road fatalities in 2022-23. Pedestrians accounted for 86% of these reported deaths while males made up 77% of the reported fatalities (Addis Ababa Police Commission et al., 2024).

A key cause of these safety issues is the prevalence of informal minibus systems, which facilitate the majority of public transit in Addis Ababa (and Sub-Saharan Africa more broadly). Most minibus drivers do not own the vehicles they drive, but instead work under 'target contract' arrangements, where the bus owner sets a daily revenue target for the driver to meet, and the driver keeps the residual revenue. This means drivers are incentivized to be reckless in order to maximize their daily earnings. For instance, although minibuses only account for 10% of all motor vehicles circulating in Kenya, they are responsible for 70% of all accidents (Macharia et al., 2009). Likewise, in Addis Ababa, minibuses account for a quarter of all crashes causing injuries, and 48% of minibuses were caught speeding in 2023 (Addis Ababa Police Commission et al., 2024).

The Addis Ababa city administration has introduced various interventions to reduce accidents including safer street designs, enforcement of traffic laws, and mass media campaigns (Addis Ababa Police Commission et al., 2024). Despite these efforts, RTAs remain high (Baru et al., 2019). As a result, policy makers in Ethiopia (especially the Addis Ababa Transport Bureau and Addis Ababa Traffic Management Authority) are keen to obtain rigorous evidence on the effectiveness of alternative monetary and/or behavioral interventions in improving driving behavior and road traffic accidents in Addis Ababa.

Based on this policymaker demand, our study runs a randomized controlled trial (RCT) to test the effectiveness of two novel and feasible interventions/programs, one financial and one non-financial, in improving the safety behaviour of minibus drivers in Addis Ababa.

Specifically, our **research question** is – How effective are monetary incentives vs non-financial/behavioural incentives at improving driver behaviour and urban road safety?

Our RCT comprises three study arms: one control group and two treatment groups. First, a GPS monitoring device was installed in all vehicles selected for the study, which recorded location and driving behavior—including instances of speeding (which is defined as going over 70 km/hr) and harsh braking at 5-second intervals. Based on this information, we were able to compute the 'safety score'/number of safety alerts triggered by each driver, defined as the sum of their speeding and harsh braking alerts they triggered over a particular time period (day or week).

Drivers assigned to the control group had a GPS device installed in their vehicles; however, neither the drivers nor the vehicle owners received any information regarding recorded driving behavior during the study period. For drivers in the non-financial treatment group, their weekly safety scores were shared publicly within a Telegram group composed of participating minibus drivers, thereby introducing peer visibility and reputational considerations. This public disclosure occurred once per week throughout the six week intervention period. Drivers in the financial treatment group were provisionally allocated a maximum bonus of 1000 birr at the beginning of each week. For each recorded safety violation—such as speeding or harsh braking—a fixed amount of 100 birr was deducted from this bonus. Consequently, the final weekly payment depended directly on the driver's observed safety performance.Both of the treatment groups also received a midweek private text message telling them how many alerts they triggered so far that week.

As a part of the experiment, we also elicit the drivers' willingness to pay for the two different treatment programs, as well as their beliefs about how safely they think they would drive under the different programs. These elicitations are done at baseline, midline and endline to see how willingness to pay and beliefs about safety performance evolve over time as drivers are exposed to the different programs.

Our pilot study finds some suggestive evidence that the non-financial treatment group reduces the number of safety alerts triggered by drivers, by about 1-2 alerts on average per week. The results of the financial incentive program are less conclusive/robust, with some regression specifications suggesting a reduction of a similar size, but others suggesting no effects on safety. We also find that the willingness to pay for the financial program is higher than for the non-financial program, and this persists from baseline to endline. Finally, we find that drivers are somewhat overconfident in their ability at baseline, and become more confident by endline (i.e. they believe they will generate fewer alerts). This last pattern holds true for those in both the financial and non-financial treatment groups.

The pilot study was conducted in close collaboration with key regulatory institutions, including the Addis Ababa Traffic Management Authority and the Addis Ababa Road and Transport Bureau, both of which have expressed strong interest in identifying scalable and cost-effective strategies to improve road safety in the city. Engagement with these institutions ensured that the interventions were designed to be operationally feasible within the existing regulatory framework. In particular, we worked closely with the Addis Ababa Road and Transport Bureau to facilitate the participation of minibus owners and to align the study design with ongoing policy priorities. As such, the findings are intended not only to contribute to the academic literature on incentives and behavioral change, but also to inform the design of practical road safety policies in Addis Ababa.

3. Literature Review

Our study related to two distinct literatures in economics. First, our study relates to a literature on how monitoring and financial incentives impact safe driving behavior. Several papers in this literature focus specifically on minibus drivers, as in our study, but these studies also extend to drivers of personal vehicles, taxi drivers, and truck drivers. Second, our study connects to a literature that compares the effectiveness of non-financial incentives to financial incentives, particularly for workers.

In the first literature on safe driving incentives, existing research on minibus drivers has primarily focused on either improving consumer information or the ability of owners to monitor their drivers as ways to reduce reckless driving and accidents. Kelley, Lane and Schönholzer (2024) install high-frequency monitoring devices in Kenyan commuter minibuses and share the data with bus owners. They find that drivers increase their effort (number of hours worked per day) and reduce unsafe practices (e.g. speeding) in response to this intervention. In a follow-up study with the same team of authors, Lane, Schönholzer, and Kelley (2025) provide passengers with pamphlets indicating which minibus companies on long-range routes are the safest. They find significant switching of passengers towards these companies when information is displayed publicly and improvements in safety by some of the companies. In a related study, Habyarimana and Jack (2011) randomly place stickers inside Kenyan minibuses that encourage passengers to speak up against unsafe driving practices, and find significant reductions in both insurance claims and average maximum speeds.

Prior research on the drivers of personal vehicles, taxi drivers, and truck drivers has found an important role for financial incentives in shaping safe driving behavior. Perhaps the most closely related study to ours is that of Raisaro (2024), which studies the effects of financial incentives on motorbike tax drivers in Uganda who view speeding as admirable. He finds that financial incentives reduce the social image costs of defying norms in favor of speeding, thereby amplifying safe driving behavior. Drivers of personal vehicles can often receive discounts from their auto insurance company in exchange for installing a monitoring device and improving their safe driving behavior. Bolderdijk et al. (2011) partner with five Dutch insurance companies to offer young drivers insurance discounts in exchange for keeping within legal speed limits. They find significant reductions in speeding violations. Similarly, Jin and Vasserman (2026) study the first major voluntary monitoring program offered by a U.S. auto insurer in exchange for discounts and find a 30% reduction in claims by those in the program. They also find that those who opt in are advantageously selected—i.e., they are safer than the average driver. Studies using data from truck driving companies instead focus on how higher wages impact safe driving behavior. In this context, drivers are not paid for non-driving work time, so low-paid drivers may reduce breaks and work excessive hours to meet earnings targets. Rodriguez, Targa, and Belzer (2006) find that a nearly 40% increase in wages at one truck company led to a decrease in crashes, and Ju and Belzer (2004) use a panel of almost 14,000 intrastate trucking companies in the U.S. to show that higher wages are associated with fewer crashes. These studies show how drivers exhibit optimizing behavior in their safety decisions, responding to financial incentives, which reveal a willingness to accept monetary compensation for safety improvements.

In the second literature on the impacts of non-financial incentives, prior studies have shown that non-financial incentives are effective in a variety of settings for increasing worker output. Ashraf, Bandiera, and Jack (2014) partner with an HIV prevention organization in Lusaka to randomize agents selling contraceptives into financial and non-financial incentive programs. They find that their non-financial incentive—a thermometer with star stamps representing sales, whereby those who hit a sales target get a certificate at a public ceremony—are more effective than financial incentives in increasing sales. In a study of salespeople at a software vendor, Larkin (2019) finds that salespeople are willing to forgo significant earnings in commissions to qualify for “Sales Club” membership, which consists of a company-wide email, a star on their business card, and a $2,000 trip to a resort. The average salesperson in his sample is willing to forgo $27,000 for Sales Club membership, showcasing the power of non-financial incentives. In a series of experiments at a gym and online, Butera et al. (2022) find that individuals are willing to pay for social recognition and that in certain social contexts, non-financial incentives shift behavior—either gym visits or performance on a monotonous task—by similar magnitudes to financial incentives. Instead focusing on peer effects, Mas and Moretti (2009) find, using data from a large supermarket chain, that the productivity of cashiers responds strongly to whether the peer cashier observing them work is of high or low productivity. The studies reveal the potential for positively influencing the safety behavior of drivers through non-financial incentives.

4. Methodology

We ran a pilot experiment in Addis Ababa with a sample of 50 minibuses, between September and November 2025. In order to recruit minibuses into the sample, we obtained the contact details for the full list of minibus *owners* in the city from the Addis Ababa transport bureau. From this list, we randomly selected owners to call and recruit. Among those owners who agreed to participate, we then contacted the driver of their minibus and separately obtained their consent to participate in our experiment. Our sample was restricted to minibuses that were only driven by one driver, and operated under a target contract arrangement (driver pays a fixed daily fee to the owner, and keeps any residual revenue).

After recruitment, the timeline of the experiment can be seen below. We first installed GPS monitoring devices (procured from [Mella Communication Technology](#)), track precise location, speed, sharp acceleration and sharp braking at 5 second intervals. We monitored drivers for one week, to get a 'baseline' safety estimate for all drivers in the sample. We then ran baseline surveys of both drivers and owners, followed by a midline survey of drivers, and an endline survey of both drivers and owners. All surveys were conducted in Amharic by a team of 10 enumerators, who were trained extensively by the research team prior to survey deployment. Two weeks after the endline surveys, we stopped monitoring drivers and gave owners access to the devices.

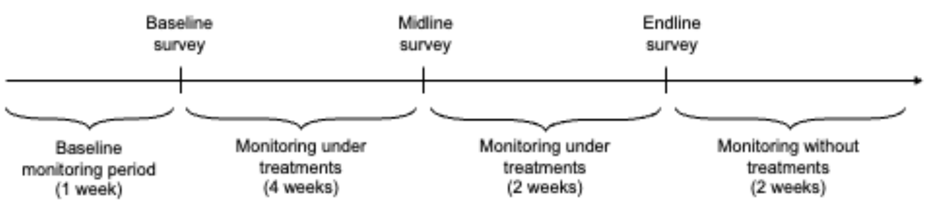


Figure 1: Timeline of pilot study

Prior to beginning the experiment, we explained to drivers the three interventions/programs that they could be assigned to. Those assigned to the control group would have their driving monitored by the researchers for the duration of the study, but information would not be shared with anyone else (including minibus owners). Those in the financial group could receive a maximum weekly bonus of 1000 birr, with 100 birr deducted for every safety alert they triggered. Those in the non-financial group would be placed into a group Telegram chat, and at the end of every week, their safety rankings (the number of safety alerts they triggered) would be publicly shared on the group. A safety alert was triggered any time the monitoring device recorded the driver as suddenly breaking, or speeding over 70 km/hr. Both of the treatment groups also received a midweek private text message telling them how many alerts they triggered so far that week.

The baseline driver survey consisted of standard socioeconomic and demographic questions (age, highest level of schooling, weekly earnings, savings/debt, religiosity etc.) We also asked how long they had been driving their minibus, and their employment tenure with their current employer. In order to gain an insight into how driver beliefs about their own safety, we asked how many alerts they think they generated during our one week baseline monitoring period, and how many alerts they believe they *would* generate under the control group, financial treatment, and non-financial treatment, respectively.

Lastly, we wanted to elicit drivers' willingness to pay (WTP), at baseline, for the two different treatment programs. In order to do this, we used a coarse multiple price list (MPL) method. Essentially, we ask drivers to suppose they are in the control group to begin with. Then, for both the financial and non-financial programs, we ask them whether they would prefer to be in that program, or to remain in the control group plus a certain amount of money. An example of this for the financial program can be seen in Table 1 below. The point at which the respondent switches from choosing the program to choosing the money, gives us a (coarse) estimate of their WTP for the program. E.g. if a driver selected 'switch program' until part c, and then switched to 'money for sure' in part d, then we know their WTP is somewhere between 1000-1500 birr. In order to ensure that respondents understood the MPL procedure, we had them do a practice round using a bar of soap as the item under consideration.

In order to make these WTP elicitations incentive compatible, we told the respondents that after they fill out the MPL, there was a possibility that their choices would determine which program they were assigned to (instead of being randomly assigned). This mechanism was implemented as follows. Before the survey, we gave respondents a sealed envelope with an amount written in it. After completing the survey, respondents drew a number from 1-6 out of a hat. If the number was a 1, then they opened the envelope and the choice that they made for the financial program that corresponded to that amount, was the choice that was implemented for them. For example, if they drew a 1 and the envelope amount was 1000, they would look in row c (third row) of Table 1. If they chose to switch to the program, then they would be assigned to the program, but if they chose the money, they would receive 1000 birr and stay in the control group.

If they drew the number 2 from the hat, the same process was followed, but we looked at their MPL for the non-financial program instead. If they drew the numbers 3 to 6, then their answers didn't matter and they were just assigned a program randomly. *So, in practice, only about ⅔ of the 50 drivers in our sample were actually randomly assigned to a program*. Therefore, the treatment effect estimation we do in the next section will only look at this randomly assigned sample.

Which do you prefer?	Switch program	Money for sure [GPS only]
a. Get 2000 birr and lose 100 birr for each safety alert **OR** get 0 birr for sure?		
b. Get 2000 birr and lose 100 birr for each safety alert **OR** get 500 birr for sure?		
c. Get 2000 birr and lose 100 birr for each safety alert **OR** get 1000 birr for sure?		
d. Get 2000 birr and lose 100 birr for each safety alert **OR** get 1500 birr for sure?		
e. Get 2000 birr and lose 100 birr for each safety alert **OR** get 2000 birr for sure?		

Table 1: Multiple price list survey for financial program

Which do you prefer?	Soap	Money
a. Soap **OR** 0 birr?		
b. Soap **OR** 50 birr?		
c. Soap **OR** 100 birr?		
d. Soap **OR** 150 birr?		
e. Soap **OR** 200 birr?		

Table 2: Multiple price list practice exercise

The midline and endline driver surveys also conducted similar WTP elicitations to see if drivers' preferences for the different programs changed over time as they were exposed to the programs. In order to make these incentive compatible, we allowed a randomly selected subset of drivers to switch programs midway through the study, based on their answers to the MPL. We also exclude these drivers from the treatment effect estimation.

The baseline and endline owner surveys were much more succinct. We asked basic demographic and socioeconomic questions, as well as how old their bus is, how many other buses they own, and how many drivers they employ. On the revenue side, we ask what daily revenue target they set, how often they adjust the revenue target, and how often they plan to look at the data from the monitoring device once we give it to them (after the study is complete).

In terms of analysis methods, as this is a pilot study with a very limited sample, we compute various basic summary statistics and data visualizations (e.g. scatter plots and histograms) to understand the broad patterns in the data. We also utilize the high frequency nature of our outcome data (safety alerts) to run panel regressions at both the daily and weekly level, as well as a cross-sectional regression of the two treatments on the average change in weekly alerts from the pre-treatment to post-treatment period.

Specifically, for the panel regressions we compute the following specification, where Y_{it} is the number of alerts generated in either a day or a week, $FinTreat_i$ and $NonFinTreat_i$ are indicator variables taking value 1 if the driver is that particular program, or zero otherwise, $BLSafety_i$ is the level of baseline (pre-treatment) safety of a driver, and γ_t are either date or week fixed effects.

$$Y_{it} = \alpha + \beta_1 FinTreat_i + \beta_2 NonFinTreat_i + \delta BLSafety_i + \gamma_t + \epsilon_{it}$$

The cross-sectional regression takes a simpler form, where Y_{it} is now the average change in weekly alerts between the pre and post period, and there are naturally no time fixed effects.

$$Y_i = \alpha + \beta_1 FinTreat_i + \beta_2 NonFinTreat_i + \delta BLSafety_i + \epsilon_i$$

5. Results

5.1 Summary statistics

	Observations	Median	Mean	SD	Min	Max
Baseline WTP (fin.)	47	1500.00	1340.43	752.58	-500.00	2000.00
Baseline WTP (non-fin.)	48	1250.00	904.17	1136.80	-1000.00	2000.00
Age	49	34.00	35.31	8.46	24.00	73.00
Driving Experience (yrs)	49	8.00	9.73	7.55	1.00	45.00
Daily Rev. Target	49	1500.00	1389.80	259.20	1000.00	2200.00
Weekly Expenses	49	4000.00	4308.49	980.71	2600.00	7000.00
Savings	49	6000.00	21122.45	46734.01	0.00	250000.00
Debt	49	0.00	14877.55	59322.44	0.00	400000.00
Forecasted alerts (control)	49	4.00	4.57	4.83	0.00	28.00
Forecasted alerts (fin.)	49	2.00	2.82	2.46	0.00	12.00
Forecasted alerts (non-fin.)	49	2.00	2.47	2.35	0.00	12.00
Weekly Alerts (baseline)	49	1.00	13.89	51.43	0.00	293.50
Weekly Alerts (treatment)	49	1.33	8.37	34.33	0.00	239.67
Completed High School	49	1.00	0.63	0.49	0.00	1.00
Very Religious	49	1.00	0.84	0.37	0.00	1.00
Daily Take-Home Rev.	49	3600.00	3180.61	1820.11	500.00	6550.00

Table 3: Baseline summary statistics of drivers

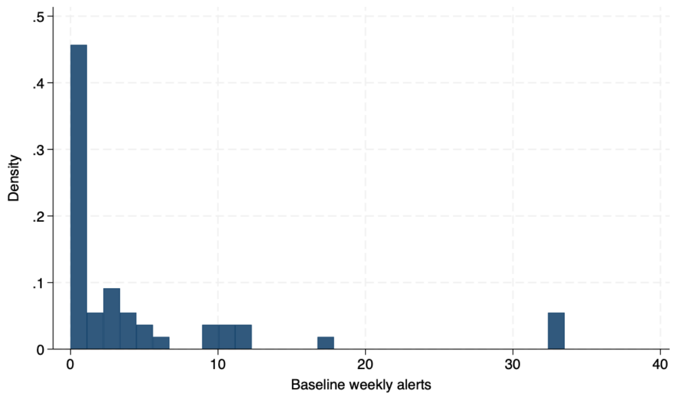
	Observations	Median	Mean	SD	Min	Max
Age	47	46.00	46.62	9.25	30.00	69.00
No. of Minibuses Owned	47	1.00	1.11	0.31	1.00	2.00
Years Owning Minibus	47	4.00	5.40	3.40	1.00	13.00
Daily Revenue	47	1300.00	1261.70	280.96	700.00	2000.00
Completed High School	47	1.00	0.79	0.41	0.00	1.00
Age of Minibus	47	24.00	23.83	5.26	4.00	38.00
Rarely Adjusts Rev. Target	47	1.00	0.83	0.38	0.00	1.00

Table 4: Baseline summary statistics of owners

We first present baseline summary statistics of drivers and owners to characterize broad patterns in our sample. Drivers are middle-aged and have a significant amount of experience (9 years on average) driving minibuses, suggesting inexperience is unlikely to be driving unsafe behaviours.

Moreover, they have high levels of savings, minimal debt and high take-home revenues, suggesting that the target contract arrangement is not as stress-inducing as initially thought. 84% of our driver sample is also very religious, and 63% have completed high school or beyond. In terms of baseline safety, we see a median number of weekly alerts of just 1, suggesting that drivers are safer than expected in this sample, albeit with a right tail of some very unsafe drivers.

Owners are also middle-aged (but older on average than drivers), and also well-educated. Most own a single minibus, and own this vehicle for a long time (an average of 5 years). Most owners (83% of our sample) rarely adjust daily revenue targets for their drivers.



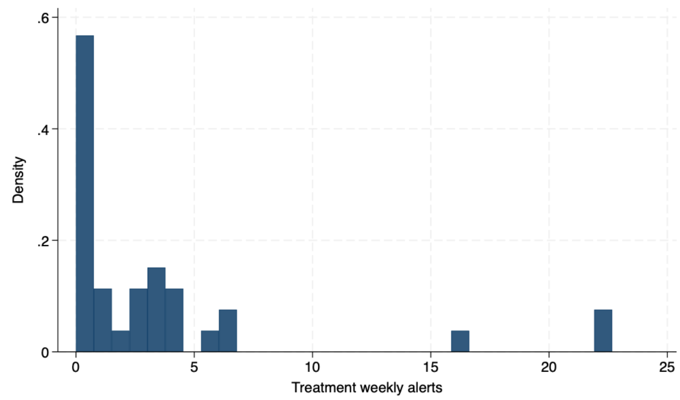


Figure 2: Histograms of weekly alerts at baseline (left) and during treatment period (right)

The histograms above provide a more detailed insight into the baseline safety of drivers, again highlighting that for the most part our sample is very safe at baseline. A visual comparison of the left and right histograms suggests that there does seem to be a slight leftward shift in the distribution of weekly alerts in the treatment period, indicating greater safety/caution by drivers.

5.2 Treatment effects

	Prob. of alert	Daily alerts	Weekly alerts	Weekly alerts change
Fin. Incentive	-0.049	-0.126	-0.491	-1.632
	(0.070)	(0.282)	(2.088)	(1.029)
Non-Fin. Incentive	-0.084	-0.324	-2.370	-0.963
	(0.053)	(0.241)	(1.720)	(0.703)
Control Mean	0.29	0.61	4.43	0.75
Observations	2100	2100	300	49

Table 5: Treatment effects of programs (choice group included)

	Prob. of alert	Daily alerts	Weekly alerts	Weekly alerts change
Fin. Incentive	-0.028	0.221	1.579	-1.167
	(0.120)	(0.316)	(2.336)	(1.543)
Non-Fin. Incentive	-0.067	-0.029	-0.151	-1.704
	(0.097)	(0.191)	(1.425)	(1.295)
Control Mean	0.21	0.30	2.17	0.10
Observations	882	882	126	21

Table 6: Treatment effects of programs (choice group excluded)

The two tables above show the results of the regression specifications outlined at the end of section 3, which aim to obtain average treatment effects of both the financial and non-financial programs. Columns 1 and 2 are daily panel regressions – in column 1 the outcome variable is an indicator taking value 1 if at least one alert was generated that day, in column 2 the outcome is the number of daily alerts. Both of these specifications have date fixed effects, and control for the number of baseline alerts generated. Column 3 is a panel regression, but the outcome is now aggregated to the weekly level (number of alerts generated per week, week fixed effects). Column 4 is a cross-sectional regression, with the outcome being the difference in average weekly alerts in the post treatment vs pre-treatment period.

While we present results for the full sample (Table 5), this also includes those who were able to choose which treatment group to go into (who we call the 'the choice group') because of their WTP elicitation at either baseline or midline. Thus, we prefer to use the smaller, but truly randomized sample, in Table 6.

We find suggestive evidence that the non-financial treatment reduced alerts, in both the panel and cross-sectional specifications. On the other hand, while the financial program reduced alerts in the cross-sectional specification, this result is not as robust as the sign flips in the panel regressions. The magnitudes of the decreases in both groups are quite small, with the cross-sectional regression suggesting an average decrease of 1-2 alerts per week between the post and pre-treatment periods, for those in the non-financial program (relative to those in the control group).

In general, these results should be viewed as very preliminary and suggestive, as the sample size for the pilot study was very small, meaning standard errors across all specifications are very large.

5.3 Baseline driver beliefs

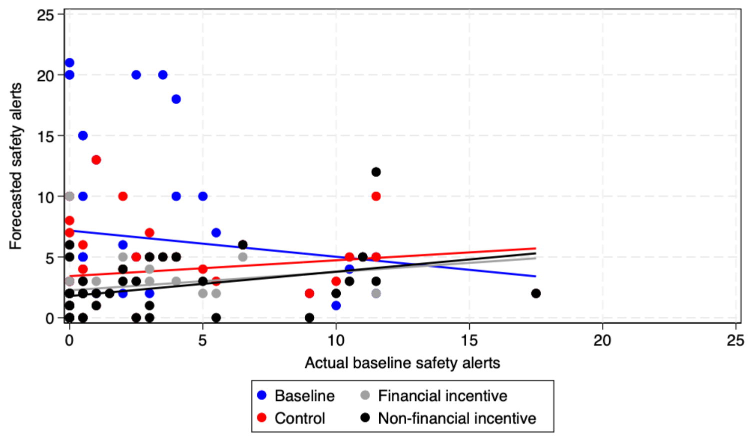


Figure 3: Baseline driver beliefs about past driving, and future driving under different programs

The above figure shows a few interesting patterns regarding driver beliefs at baseline. The blue scatter plot shows the correlation between drivers actual safety at baseline vs how they thought they performed at baseline when asked. We see a slight negative, but essentially flat, relationship between these two variables, suggesting that drivers do not really know how safe they are. After asking them about their perceived baseline safety, we then tell them how they actually did, and then ask them three more questions: how well do you think you would do under our three programs (control, financial, non-financial)? The results of those correlations can be seen in the grey, red and black scatter plots. Here we see some evidence of overconfidence, drivers consistently say they will do better than what they have actually done at baseline, irrespective of which program they are asked to consider.

5.4 Willingness to pay at baseline

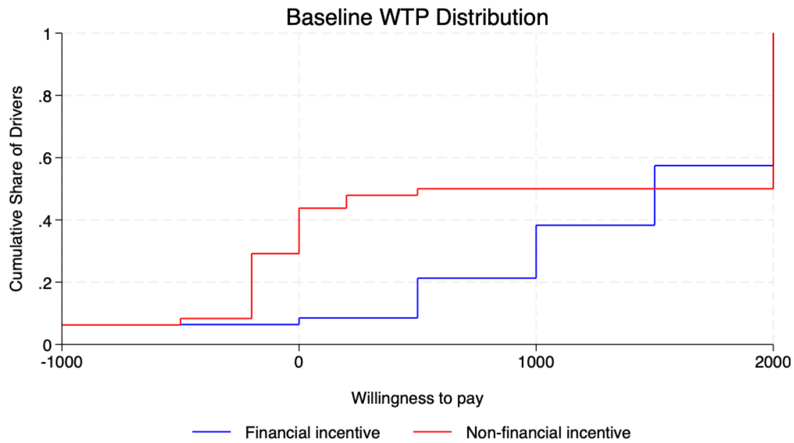


Figure 4: Distribution of WTP for both programs at baseline

	Fin. WTP	Non-Fin. WTP
Forecasted alerts (Fin.)	-33.437	
	(37.430)	
Forecasted alerts (Non-Fin.)		-87.471
		(62.029)
Observations	47	48

Table 7: Correlation between forecasted alerts and WTP at baseline

Figure 4 shows that the WTP at baseline is generally higher for the financial program than the non-financial program (except for at very high values above 1500 birr). Moreover, there is clearly variation in WTP for each program (i.e. it is not as if everyone only wants to pay either -1000 birr or 2000 birr for the programs). Table 7 shows that our survey instrument was successfully understood by respondents, because a higher number of forecasted alerts for a particular program is correlated with a lower WTP for that program (which is what one would logically expect).

5.5 Changes in WTP over time

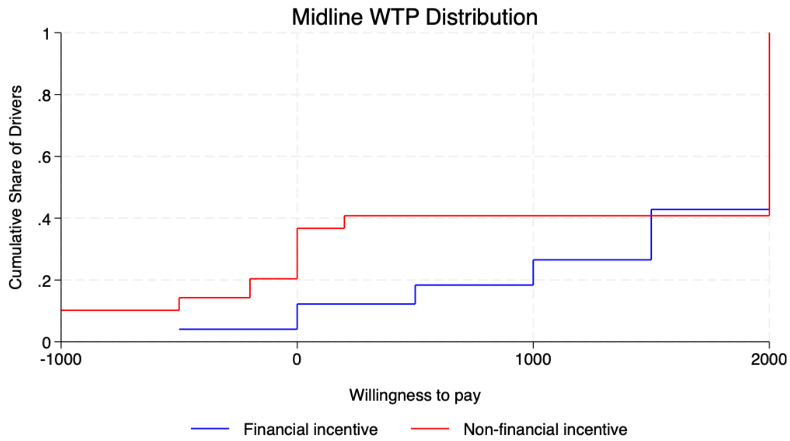


Figure 5: Distribution of WTP for both programs at midline

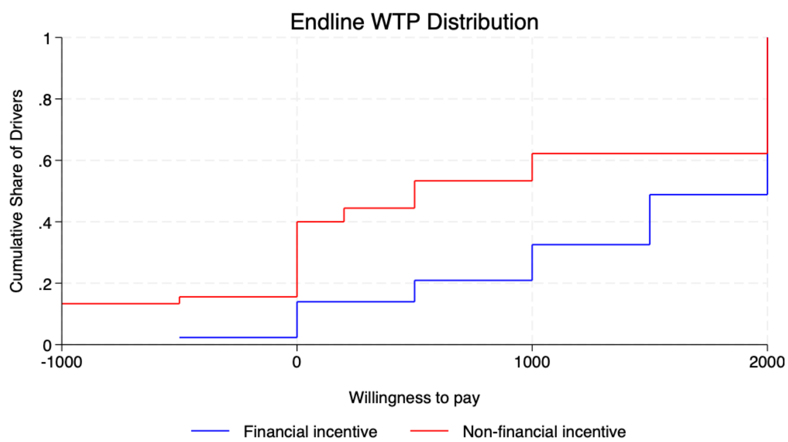


Figure 6: Distribution of WTP for both programs at endline

Comparing Figures 5 and 6 to Figure 4, we find that WTP for the financial program remains higher than WTP for the non-financial program at both midline and endline. Although we see some convergence between baseline and midline between the two groups, they diverge again at endline. In general, there are not considerable movements or changes in patterns of WTP over time.

5.6 Changes in beliefs over time

	Change in beliefs	Change in beliefs
Fin. Incentive	0.196	
	(0.828)	
Non-Fin. Incentive	-1.105	
	(0.816)	
Change in alerts		0.025
		(0.081)
Observations	47	47

Table 8: Correlations between programs and change in beliefs between baseline and midline

	Change in beliefs	Change in beliefs
Fin. Incentive	-0.583	
	(1.570)	
Non-Fin. Incentive	-2.021	
	(1.570)	
Change in alerts		0.176
		(0.162)
Observations	44	44

Table 9: Correlations between programs and change in beliefs between baseline and endline

Finally, we study how changes in beliefs between baseline and midline, and baseline and endline, correlate with participation in our two treatments. At midline, we find that those in the financial program believe that they will generate more alerts (i.e. think they are more unsafe) than at baseline. The opposite is true for those in the non-financial program, who believe they will now generate fewer alerts. At endline, those in both treatment groups think they will be safer on average, than they were at baseline (i.e. they update their beliefs to be more confident in their driving). Moreover, those with a higher treatment effect (a more negative change in alerts) seem to also become more confident over time.

6. Conclusion

In summary, we ran a pilot study with 50 minibus drivers in Addis Ababa to determine the effectiveness of financial vs non-financial incentives in improving their road safety behaviour. In order to effectively measure road safety behaviour, we installed GPS monitoring devices in each minibus that gave us very high-frequency information on when drivers were speeding (going over 70 km/hr) and/or suddenly braking.

Drivers were randomized into one of three groups: (1) a control group who were simply monitored by the researchers, (2) a financial treatment group who received a maximum bonus payment of 1000 birr per week, with 100 birr deducted for every speeding or sudden braking violation detected by the device, and (3) a non-financial treatment group whose safety rankings were revealed in a Telegram group chat at the end of every week. Both treatment groups also received a mid-week private text message with information on how many alerts they had generated thus far.

We found suggestive evidence of the non-financial treatment being effective in improving driver safety, although the magnitude of the effects are small. On the other hand, the financial incentive does not seem to show much impact on safety behaviour. We also conducted willingness to pay and belief elicitations of all drivers at baseline, midline and endline. We found consistently higher WTP for the financial treatment, and we found that drivers across both treatment groups believe themselves to be safer at endline than at baseline.

There are several limitations of this study. Firstly, the sample size is very small (50 drivers, reducing to fewer than 30 once we account for those who were able to choose their program), meaning statistical inference is very challenging and the standard errors on all results are high. Secondly, recruitment in this setting was quite challenging, as many drivers were reluctant to participate in a program that tracked them. As such, we believe we have a sample that is positively selected in terms of being safer than the average minibus driver in Addis Ababa. This could explain the surprisingly small number of weekly alerts we see at baseline, and means that our results are likely a lower bound on the effects you would see if you ran this program on 'average' minibus drivers. Finally, due to budget constraints, the financial incentive offered to the drivers was (in hindsight) somewhat small relative to their weekly earnings, which could explain the lack of effects observed for the financial program.

Future research could focus on scaling up this pilot program to a larger sample size of around 250-300 minibus drivers, as in similar papers in this literature such as Kelley, Lane and Schonholzer (2024). This would enable better statistical inference. Future work could also do a similar study with rideshare workers, who are an important and growing population of drivers in African cities.

7. Bibliography

Abegaz, T., & Gebremedhin, S. (2019). Magnitude of road traffic accident related injuries and fatalities in Ethiopia. PLOS ONE, 14(1), e0202240. https://doi.org/10.1371/journal.pone.0202240

Addis Ababa City Administration Transport Bureau (AATB). (2023). Addis Ababa annual road safety report. Retrieved Feb 10, 2026, from https://www.Aatb.gov.et/download/annual-reports/

Addis Ababa Police Commission, Addis Ababa Traffic Management Authority, & Bloomberg Philanthropies Initiative for Global Road Safety. (2024). Addis Ababa road safety report 2022–2023.

Ashraf, N., Bandiera, O., & Jack, B. K. (2014). No margin, no mission? A field experiment on incentives for public service delivery. Journal of Public Economics, 120, 1–17.

Baru, A., Azazh, A., & Beza, L. (2019). Injury severity levels and associated factors among road traffic collision victims referred to emergency departments of selected public hospitals in Addis Ababa, Ethiopia: The study based on the Haddon matrix. BMC Emergency Medicine, 19(1), 2. https://doi.org/10.1186/s12873-018-0206-1

Bolderdijk, J. W., Knockaert, J., Steg, E.M., & Verhoef, E. T. (2011). Effects of pay-as-you-drive vehicle insurance on young drivers' speed choice: Results of a Dutch field experiment. Accident Analysis & Prevention, 43(3), 1181–1186.

Butera, L., Metcalfe, R., Morrison, W., & Taubinsky, D. (2022). Measuring the Welfare Effects of Shame and Pride. American Economic Review, 112(1), 122–168.

Deme, D. (2019). Road Traffic Accident in Ethiopia from 2007/08–2017/18 (Since Ethiopia Millennium). https://sryahwapublications.com/article/download/2637-5796.0202003

Habyarimana, J., & Jack, W. (2011). Heckle and chide: Results of a randomized road safety intervention in Kenya. Journal of Public Economics, 95(11–12), 1438–1446.

Jin, G. Z., & Vasserman, S. (2026). Buying data from consumers: The impact of monitoring in U.S. auto insurance. NBER Working Paper No. 29096.

Ju, S., & Belzer, M. H. (2004). Follow the money: Trucker pay incentives, working time, and safety. The Economic and Labour Relations Review, 35(1), 7–26.

Kelley, E. M., Lane, G., & Schönholzer, D. (2024). Monitoring in small firms: Experimental evidence from Kenyan public transit. American Economic Review, 114(10), 3119–3160.

Lane, G., Schönholzer, D., & Kelley, E. (2025). Information and strategy in lemon markets: Improving safety in informal transit. Working paper.

Larkin, I. (2019). Paying $30,000 for a gold star: An empirical investigation into the value of peer recognition to software salespeople. Working paper.

Macharia, W. M., Njeru, E. K., Muli-Musiime, F., & Nantulya, V. (2009). Severe road traffic injuries in Kenya: Quality of care and access. African Health Sciences, 9.

Raisaro, C. (2024). Incentives Justifying Nonconformity: Experimental Evidence from Motortaxi Organizations in Uganda. Working paper.

Rodriguez, D. A., Targa, F., & Belzer, M. H. (2006). Pay incentives and truck driver safety: a case study. ILR Review, 59(2), 205–225.

World Health Organization. (2026). Road traffic injuries. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries