Results of a large-scale randomized behavior change intervention on road safety in Kenya

James Habyarimana[a,1] and William Jack[b,1]

[a]McCourt School of Public Policy, Georgetown University, Washington, DC 20057; and [b]Department of Economics, Georgetown University, Washington, DC 20057

Edited by Kenneth W. Wachter, University of California, Berkeley, CA, and approved July 6, 2015 (received for review November 17, 2014)

Road accidents kill 1.3 million people each year, most in the developing world. We test the efficacy of evocative messages, delivered on stickers placed inside Kenyan matatus, or minibuses, in reducing road accidents. We randomize the intervention, which nudges passengers to complain to their drivers directly, across 12,000 vehicles and find that on average it reduces insurance claims rates of matatus by between one-quarter and one-third and is associated with 140 fewer road accidents per year than predicted. Messages promoting collective action are especially effective, and evocative images are an important motivator. Average maximum speeds and average moving speeds are 1–2 km/h lower in vehicles assigned to treatment. We cannot reject the null hypothesis of no placebo effect. We were unable to discern any impact of a complementary radio campaign on insurance claims. Finally, the sticker intervention is inexpensive: we estimate the cost-effectiveness of the most impactful stickers to be between \$10 and \$45 per disability-adjusted life-year saved.

road safety | governance | accountability | consumer empowerment

Road accidents represent a large and growing cause of death in the developing world (1). About 1.3 million people are estimated to die each year on the road, and between 20 and 50 million are injured. By 2030, road accidents will be the fifth leading cause of death worldwide, killing more people than HIV/AIDS, tuberculosis, and malaria combined. Among 15–29 y olds, road accidents are already the single largest cause of death. In light of these figures, the World Health Organization (WHO) has called for urgent action to address the epidemic of road injuries and deaths.

In this paper, we report results of a scaled-up road safety experiment in Kenya between 2011 and 2013, called Zusha! (Swahili for "Protest!"), which was aimed at reducing accidents involving the country's 14-seater minibuses, or matatus. Our intervention was aimed at promoting agency among matatu passengers, empowering them to challenge the driver to slow down and drive less recklessly if they felt their safety was compromised. To this end, we posted stickers with evocative messages inside the vehicles, exhorting the passengers to act, with phrases like "Don't let a reckless driver get away with murder," and others. We tested the efficacy of different kinds of messages, including the role of images in general in eliciting a response, and the relative merits of using fear vs. reason to motivate behavior change. We also present evidence that messages promoting collective action were more effective than those aimed simply at alleviating individual-level constraints. That is, inaction appears to be associated more with coordination failure than with a lack of information as such.

Overall, the stickers reduce insurance claims of matatus assigned to treatment groups by between one-quarter and one-third on an intent-to-treat (ITT) basis. Among the roughly 8,000 vehicles in the treatment groups, the reduction was 25%, and we estimate that about 140 accidents were avoided per year, and about 55 lives were saved annually. Using data on the age distribution of likely passengers derived from a survey we conducted, we conservatively estimate the cost-effectiveness of the intervention as being between \$13 and \$60 per disability-adjusted life-year (DALY) saved. The cost-effectiveness of the more impactful stickers, which reduced claims by 34%, was between \$10 and \$45 per DALY saved.

In addition to the sticker intervention, a radio campaign was aired on two local radio stations that cover certain areas around Mt. Kenya in the central highlands and in the capital city Nairobi. The campaign was activated on a weekly basis on five occasions over the course of the first 6 mo of 2012. We examine insurance claims for events that took place over this period in areas with radio coverage and without, but in contrast to the large effects of the stickers, find no significant impacts.

In Habyarimana and Jack (2), we documented the results of a small pilot study of a similar intervention with a single treatment group and a control. The treatment included a bundle of five stickers, three with text only and two with grisly images of injuries, as well as a weekly lottery that drivers of treated vehicles could win if they were found to be in compliance. Although compliance with the treatment was not perfect, it was sufficiently high to allow us to estimate significant impacts. Vehicles assigned to the treatment group were 50% less likely to file an insurance claim over the following 12-mo period and were 60% less likely to file a claim for an accident involving an injury or death. Using treatment assignment as an instrumental variable, we estimated a local average treatment effect on those who complied with the treatment assignment equivalent to an 80% reduction in accident rates.

The experiment reported in this article was more comprehensive, with a sample five times the size of the pilot, allowing us to incorporate fully eight treatment arms as described in detail below and a placebo. The placebo allowed us to address a potential shortcoming of the pilot: that the lottery itself could have induced better driving. Although eligibility to win the lottery was based only on the retention of the stickers and not on remaining accident free, there was a concern that the prospect of winning the lottery might itself also have induced drivers to be more

Significance

Road accidents kill 1.3 million people each year, most in the developing world. Evocative messages inside Kenyan matatus, or mini-buses, that promote passenger agency and legitimize complaints against dangerous driving are found to reduce average maximum speeds and average moving speeds by 1–2 km/h and insurance claims by between one-quarter and one-third. The cost-effectiveness of the most impactful stickers is between \$10 and \$45 per disability-adjusted life-year saved.

Author contributions: J.H. and W.J. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence may be addressed. Email: jph35@georgetown.edu or wgj@georgetown.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1422009112/-/DCSupplemental.

Table 1. Phase 1 treatment assignments and retention in phase 2

Sticker type	Phase 1 recruitment	Phase 2 retention		
Pure control	2,093	82%		
Placebo	1,759	77%		
Treatment stickers	7,885	73%		
Total	11,737	75%		
	Individual action		Collective action	
Treatment groups	Phase 1 recruitment	Phase 2 retention	Phase 1 recruitment	Phase 2 retention
Text only	971	73%	979	70%
Supportive (voice)	1,000	71%	998	72%
Consequence aversion (injuries)	974	75%	980	76%
Event aversion (crashes)	992	70%	991	72%
Total	3,937	72%	3,948	73%

careful: either because the prize made life more worth living or because they misinterpreted the eligibility rules and thought they could win the prize if they had a safe driving record. The placebo arm in our expanded study thus included a relatively neutral sticker,* retention of which gave drivers the same chance of winning the lottery as those of vehicles in the treatment arms. However, we find no evidence of a placebo effect, suggesting the treatment stickers themselves lead to a reduction in accidents.

Most other interventions aimed at inducing changes in driving behavior have been directed at the driver, and some have used shock therapy to that end. Guria and Leung (3), for example, found effects of an advertising campaign in New Zealand against speeding and drunk driving. Other studies have reported impacts of seat belt rules, speed humps, and speed cameras (1, 3–8), but few randomized evaluations, if any (apart from our pilot), have been conducted.

Theory of Change

The factors that lead to road crashes and that are within the control of individual decision makers include speed, use of attention-impairing substances, seat belt use, vehicle maintenance, and general effort or care. If all stakeholders make well-informed and well-coordinated decisions, then the incidence of road accidents will be in some sense efficient. However, at least three sources of market failure could prevail on the road. One is the obvious externality imposed by a reckless matatu driver on other road users external to his vehicle, including other drivers, passengers, and pedestrians. This source of inefficiency is not directly addressed by our intervention. A second reflects a simple lack of information on the part of passengers: the social norm may dictate that passengers do not question the driver and that complaining is simply not an option. That is, even when gains from trade between the driver on the one hand and the passengers on the other exist, bargaining might not take place. Finally, a free-rider problem could exist among passengers inside a matatu, whereby the collective preferences of passengers outweigh those of the driver but are not expressed. If there is a psychic or social cost to complaining (perhaps no one wants to be perceived as weak), but the benefits of doing so accrue to all riders, then complaining has the attributes of a public good, and social pressure will likely be undersupplied.[†]

The interventions we evaluate are aimed at correcting the second and third market failures identified above and could act either to directly inform individuals of the feasibility of complaining or to lower the cost of doing so. More indirectly, the messages—conveyed visually to matatu passengers on stickers inside the vehicle, as well as through radio announcements—could legitimize complaint, allowing riders to confidently challenge the heretofore-unquestioned authority of the driver.

As well as motivating complaints, some of the stickers include a small nudge that encourages collective action (see below). Any evidence that stickers with this component have differential (positive) effects will suggest that coordination failures within minibuses are important determinants of road accidents.

The stickers could of course influence driver behavior even in the absence of passenger responses, either directly if the driver himself takes note of the messages or indirectly if he anticipates the complaints of passengers and preemptively adjusts his driving accordingly. Indeed we do not observe statistically meaningful differences between treatment and control groups in terms of observed passenger behavior, as captured by field staff who recorded passenger behavior and driver responses on nearly 1,500 trips, although the highly qualitative nature of these data limits our ability to do so in any case. Nonetheless, such inconclusive evidence is not inconsistent with the large reduced form impacts on insurance claims that we observe.

The trips taken by our field staff do, on the other hand, yield informative data on speed. Using hand-held global positioning system (GPS) devices, we find that vehicles assigned to treatment groups exhibited lower recorded maximum and average moving speeds than those in the control or placebo arms.

Finally, our treatment is at the vehicle level. Whereas most drivers drive the same vehicle on a regular basis, there is some driver rotation both within the day and across days, so drivers who have been exposed to the treatment can end up driving untreated vehicles. More importantly, passengers nearly certainly will ride on both treated and untreated vehicles. These patterns of exposure suggest there could be large spillovers across vehicles, which would hamper our ability to observe differences in outcomes. Only if there was no driver effect, and only if the passenger impact depreciated immediately after exposure, would observed differences reflect the true underlying treatment effect of the interventions. Our measured effects are thus nearly certainly underestimates of the true impacts of the intervention.

The Interventions

The experiment consisted of a total of 10 different arms as documented in Table 1. Vehicles assigned to a control group received no stickers and were ineligible for the lottery. Vehicles

*A true placebo arm would have had no sticker whatsoever and simply the lottery. The placebo stickers we used can be interpreted as a low-, but not zero-, dose intervention.

[†]Note that not all riders will derive the same, or even positive, benefit from one person's complaints. Nonetheless, lowering the costs of voicing their preferences should improve the ability of passengers to coordinate their actions.

in the other nine groups were offered four stickers each, two with English text and two with Kiswahili. Vehicles in a placebo group were assigned stickers that read simply "Travel Well" and the Kiswahili translation "Safiri Salama."

All other vehicles were assigned across four primary treatment arms, each with two subarms. In all these treatment groups, passengers were emphatically encouraged to speak up against bad driving. One primary treatment arm contained stickers with text only and no images, whereas the other three primary treatment arms included images along with the same text. Among these three arms, the "supportive" theme included images of riders shouting at the driver, the "consequence aversion" theme included images of injured passengers, and the "event aversion" theme included images of wrecked vehicles. Fig. 1 shows one of the stickers issued under each theme.

Finally, each of the four primary treatments was crossed with a collective action intervention. In the four "individual action" subarms, no changes were made to the stickers; but in the four "collective action" subarms, a small additional piece of text was added to each sticker, reading "Umoja ni nguvu." Literally this translates as "Unity is power," but the meaning might be closer to "Together we can." The consequence aversion sticker in Fig. 1 includes this motif.

From 2011 to 2013, we worked with a Kenyan insurance company that sold coverage to about 12,000 matatus through seven sales offices throughout the country.[‡] Third party insurance is mandatory for public service vehicles in Kenya, and most vehicles purchase policies on a monthly basis. Typically, an insurance agent visits a sales office, pays for the coverage, and collects a certificate that is then posted on the inside of the windscreen as evidence of insurance. The decisions to accept, insert, and retain the four stickers were effectively taken by three parties, respectively, the vehicle owner, the insurance agent (an intermediary who purchases insurance on behalf of owners), and the driver. To encourage compliance with the intervention, all three parties were enrolled in a lottery. After recruitment, each week, 10 noncontrol group vehicles were chosen at random (from a population of roughly 10,000) and inspected. If all four stickers were observed to be in the appropriate position, the agent, owner, and driver each received 5,000 Kenyan Shillings, or about $60 (roughly a week's wage for the driver). Vehicles in the placebo group were eligible for the lottery. To boost the credibility of the lottery, during the first 2 mo of the intervention (March and April 2011), advertisements were broadcast on two radio stations encouraging the agents, owners, and drivers to accept the stickers and keep them in place and included interviews with winners.

Finally, although treatment assignment was recorded and used for the purposes of implementing the lottery, it played no role in the claims review, adjustment, or settlement process, and individual claims staff were unaware of the status of any given vehicle.

Recruitment, Data, and Empirical Strategy

Recruitment and Randomization. We recruited vehicles at the point of insurance purchase in two phases, first from March to May 2011 and then from December 2011 to April 2012. At the point of recruitment, vehicles were randomized across treatment, placebo, and control arms using a random number generated as part of the computerized sales process. The 8,797 vehicles that purchased coverage in both phase 1 and phase 2 were assigned to the same groups throughout the experiment. Phase 2 was never intended to be a new experiment but rather an attempt to replenish stickers from phase 1 due to wear and tear, so in our analysis, we use only those vehicles recruited in phase 1. Table 1 reports the numbers first recruited in each treatment arm in phase 1 and the share of each group that was retained in phase 2.

Fig. 1. Examples of stickers in the placebo arm and in each treatment group.

Data. We use three sources of data in our analysis. First, we have access to complete administrative data on insurance claims from January 2009 to December 2013 from the company with which we worked. Among the sample of vehicles we recruit, we have basic identifying information (such as the license plate), as well as the identity of vehicles drawn in the lottery each week and the number of winners.

Second, we conducted a short interview with 9,807 passengers at the termination of bus rides, in which we inquired about their experiences on the trip and their exposure to the radio campaign.

Third, we collected data from a total of 4,405 matatu trips on vehicles recruited in phase 1 taken by our field staff over the course of 1 y following recruitment. The enumerators were equipped with a GPS device that automatically measured the maximum trip speed and average moving speed and could be used to manually record the location and nature of any dangerous incidents, any associated behavior of passengers, and the response of drivers. The enumerators were instructed to ride any matatu that was currently covered by our partner insurance company (something that could be inferred easily by visual inspection), without regard to the presence or absence of any stickers. They did, however, record the number plate so we could determine treatment assignment.

The total number of vehicles in our experimental sample increased from 11,737 in phase 1 to 12,512 in phase 2, due in part to the growth in the insurer's business over the period. However,

[‡]The company sold insurance to matatus that operated on both intracity and long distance routes, although these data were not collected at the point of sale. In the pilot study (2), we recruited only long distance vehicles.

Table 2. Summary statistics

Indicator	Mean	SD	Minimum	Maximum
Claims data (n = 9,358 incidents)*				
No. of claims per incident	1.993	3.516	1	83
No. of fatalities per incident	0.105	0.49	0	21
Share of incidents with at least one fatality	0.078	0.269	0	1
No. of claims with broken bones per incident	0.155	0.517	0	11
Share of incidents with at least one injury	0.444	0.497	0	1
No. of claims with soft tissue injury per incident	1.051	3.114	0	77
Share of incidents with nonvehicle claimants	0.188	0.39	0	1
Share of incidents with vehicle damage claimant	0.256	0.436	0	1
Passenger survey (n = 9,807)†				
Share male	0.68	0.46	0	1
Age	32.69	8.38	13	83
Share speak Kikuyu	0.35	0.48	0	1
Share speak Meru	0.25	0.43	0	1
Share had at least two bus trips in past week	0.58	0.49	0	1
Share ever exposed to radio spot	0.65	0.48	0	1
Trips data‡				
Speed (n = 4,405)				
Distance covered on trip, km	69.07	50.97	3	229
Maximum speed reached, km/h	95.95	15.51	25	155
Moving average speed, km/h	45.88	17.04	3	105
Events (n = 1,471)				
Share of events reported as reckless	0.15	0.36	0	1
Share of events passengers speak to driver	0.18	0.38	0	1

*Claims data for January 1, 2009–December 13, 2013, during which period there were 9,538 incidents. An incident is a crash or other event that results in at least one insurance claim. There can be multiple claims (e.g., by different injured individuals) associated with a single incident. Data reflect all claims, not just claims by vehicles in our experimental sample.

†Kikuyu and Meru are dominant languages spoken in Central Kenya, where the two radio stations that aired the safety messages have the largest audiences.

‡Enumerators collected speed data automatically on GPS units; 4,405 valid records were submitted. However, events such as excessive speed, reckless driving, and passenger complaints to drivers were recorded manually. Of these, 1,471 events could be properly matched to vehicles.

this net increase reflected exit of 2,943 vehicles (25%) and entry of 3,718 new vehicles (30% of the phase 2 sample), a rate of churn that is common in the industry. Table 2 reports summary statistics for vehicles recruited in phase 1 from the three data sources described above.

Balance and Attrition. We collect limited information on the vehicles at recruitment, but use license plate attributes (the letters of which provide an indicator of the time since registration, but not necessarily vehicle age) and prerecruitment claims data to present a balance test in Table 3.[§] The table reports comparisons of vehicles in the control and placebo group, and any treatment group.[¶]

There is no imbalance on license plate attributes at baseline (Table 3), and if anything, as reported, the treatment group was more likely to have had a previous claim than the control and placebo groups and had higher annualized accident rates in previous years (although the number of previous claims per accident was lower in the treatment group). To the extent that these differences reflect a divergence in underlying group characteristics, a simple comparison of treatment with control and placebo vehicles would underestimate the treatment effect of the stickers. On the other hand, if the observed preintervention accidents and claims follow nonparallel trends such as a mean reversion process, then the difference in differences estimate would overstate the treatment effect. To assess this possibility we conduct a falsification test using data from earlier periods. The results, available in Table S3, show that preintervention trends were actually higher in the treatment group, suggesting an even larger treatment effect than we estimate with the simple double difference. Indeed, under the assumption that the preintervention trends (as opposed to levels) in each group would have been maintained, the estimated treatment effect of the stickers is roughly twice as large.

Finally, attrition at phase 2 differed across groups, being highest among treatment vehicles, and lowest among control vehicles (Table 3). Table S4 provides a test for balance of the nonattrited sample on preintervention characteristics. The direction and extent of imbalance is consistent with that reported in Table 3. In addition, using data from 2009 onward, we find no evidence that attrition is systematically related to insurance claims history, nor, by inference, to the potential impact of the stickers on driver behavior.[#] Nonetheless, we propose that any nonrandom attrition would be biased toward more risky drivers and would attenuate measured effects.

Fig. 2 shows the number of weekly lottery winners over the course of the two phases of the project. Compliance with assignment to sticker groups, as measured by the number of winners out of 10 drawn, was initially very high, but fell over the first 6 mo of the intervention. During the second phase, the

§We do not have comprehensive data on the insurance coverage status of the recruited vehicles over the full history of claims data to which we have access. Instead, we only see if and when those vehicles experienced an event that led to a claim.

¶Table S1 reports these balance tests for the individual action and collective action groups separately, and Table S2 reports the results for the four message types separately. Similar patterns are observed.

#In a regression of attrition from phase 1, the coefficient on annualized prerecruitment claims rates is −0.012, with a robust SE of 0.016 and R^2 of 0.00 (n = 11,737).

Table 3. Balance test

Indicator	Control	Placebo	Treatment	Control vs. placebo	Control vs. treatment	Placebo vs. treatment
Vehicle attributes						
Share with license KA...[†]	0.583	0.567	0.566	0.017	0.017	0.000
	(0.011)	(0.012)	(0.006)	(0.016)	(0.012)	(0.013)
Share with license KB...	0.416	0.433	0.433	−0.016	−0.016	0.000
	(0.011)	(0.012)	(0.006)	(0.016)	(0.012)	(0.013)
Share with license KBK-N	0.113	0.122	0.117	−0.009	−0.005	0.004
	(0.007)	(0.008)	(0.004)	(0.010)	(0.008)	(0.009)
Share with license KBP-S	0.006	0.009	0.009	−0.002	−0.003	−0.001
	(0.002)	(0.002)	(0.001)	(0.003)	(0.002)	(0.003)
Phase 1 attrition rate	0.175	0.229	0.276	−0.054***	−0.100***	−0.046***
	(0.008)	(0.010)	(0.005)	(0.013)	(0.011)	(0.012)
N	2,093	1,759	7,885	3,852	9,978	9,644
Claims data						
Any previous claim[‡]	0.152	0.170	0.181	−0.018	−0.029***	−0.011
	(0.008)	(0.009)	(0.004)	(0.012)	(0.009)	(0.010)
Any previous injury claim	0.078	0.077	0.084	0.001	−0.005	−0.007
	(0.006)	(0.006)	(0.003)	(0.009)	(0.007)	(0.007)
Any previous fatality claim	0.015	0.016	0.016	−0.001	−0.001	0.001
	(0.003)	(0.003)	(0.001)	(0.004)	(0.003)	(0.003)
Number previous claims	2.563	1.938	1.775	0.626**	0.789***	0.163
	(0.211)	(0.141)	(0.062)	(0.256)	(0.167)	(0.151)
Annualized accident rate	0.079	0.097	0.104	−0.018**	−0.025***	−0.007
	(0.005)	(0.006)	(0.003)	(0.007)	(0.006)	(0.007)
N	2,093	1,759	7,885	3,852	9,978	9,644

Table covers only phase 1 vehicles, including both those that were found in phase 2 and those that were not. SEs in parentheses. Significant differences at **5% and ***1% levels.

[†]Vehicle license plates are of the form KXX NNNX, where X represents a letter and N a number. KA... license plates were issued before KB... plates, and among KB... plates, KBK-N were issued before KBP-S.

[‡]Previous claims refer to claims recorded before recruitment.

same pattern emerged, but starting from a lower base. This lower base may have been because the radio advertisements promoting the lottery were not aired in phase 2 or because the stakeholders had updated their beliefs about the probability of being drawn.

Empirical Specifications. In our analysis of sticker impacts below, when using pre- and postintervention claims data, we calculate intent to treat estimates of a difference in difference specification of the form

$$y_{igt} = \alpha + \beta_g D_{ig} + \gamma Post_{it} + \delta_g D_{ig} \times Post_{it} + \varepsilon_{igt}, \qquad [1]$$

where D_{ig} is a dummy variable equal to 1 if vehicle i is in treatment group g, and $Post_{it}$ is a dummy equal to 1 if the observation for vehicle i is after the adoption of the intervention. The coefficient of interest, δ_g, measures the extent to which the outcome y_{igt} for vehicle i in the post period differs from the expected level, given its baseline level. The primary outcome variable we use is insurance claims, but we also model accident severity and maximum and average recorded speeds.[||] Because of the large number of experimental groups, we present results for a variety of treatment aggregations. Similarly, after confirming the absence of a placebo effect, we combine the placebo and control groups. SEs are clustered at the vehicle level.

We use our passenger survey data to assess exposure to the radio campaign, which was deployed in three regions of Kenya—Nairobi, Nakuru, and Mt. Kenya—and was orthogonal to the sticker assignment. In addition, we model weekly claims rates with the following regression:

$$y_{ijt} = \sum_{j=1}^{3} \mu_j + \sum_{w=1}^{52} \omega_w + \psi_{2012} + \sum_{j=0}^{3} \sum_{t=-52}^{51} \delta_{jt} \mu_j D_t + \varepsilon_{it}. \qquad [2]$$

Based on the sales office at which it purchased insurance coverage on the date of recruitment, each vehicle is assigned to one of four regions: j = 0, 1, 2, 3. Region 0 is Mombasa: a location far from the other three regions that were exposed to the radio campaign. y_{ijt} is an indicator variable equal to 1 if vehicle i in region j experienced a claim in week t. We estimate this equation using data from January 2011 to December 2012.**

We include week-of-the-year fixed effects (ω_w) and a fixed effect for 2012 (ψ_{2012}). The coefficients of interest are the δ_{jt}, which multiply a dummy variable D_t equal to 1 in week t of the 2-y period. A negative and significant value of δ_{jt} for weeks coinciding with or just after the weeks in which the radio spots aired would indicate a corresponding reduction in claims associated with the campaign.

Results

In this section, we report results of the sticker intervention and the radio campaign.

[||]As we only have speed data for the period after recruitment, our intent to treat estimates of sticker impacts are based on a simple (single-difference) ordinary least squares specification on assignment.

**The sample of vehicles we use includes 26,213 that were observed between March 2011 and July 2012. We assume a balanced panel of these vehicles over the full 2-y observation period. This assumption implies that all of these vehicles' accident records are observed for the entire period. However, in the context of significant turnover, any accidents covered by other insurance companies are not captured in our data. Because we don't expect such churning to be correlated with the timing of the radio campaign, measurement error is differenced out. (We show below that the analysis remains unchanged if we restrict the sample to vehicles that are observed in at least 8 of the 16 mo.)

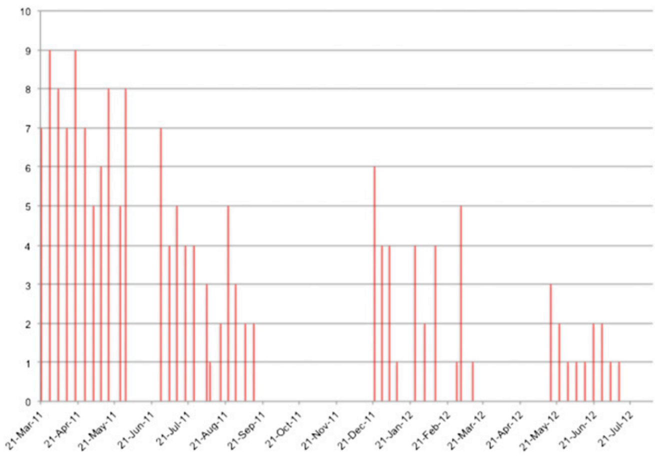

Fig. 2. Weekly lottery winners.

Sticker Intervention Results. Table 4 reports our main results of the effects of the sticker interventions, using the difference in difference specification in Eq. **1** for claims (first column), and a single difference specification for the other outcomes (second through fifth columns). For the claims regression, we exploit data spanning a period of 28 mo before the beginning of the intervention to 30 mo after. In the trips regressions, we control for the trip distance and include enumerator fixed effects. All results measure intent to treat effects. We use data on all vehicles recruited in phase 1, but disregard those that were recruited for the first time in phase 2.

Each horizontal panel presents the coefficients of regressions with treatment assignment aggregated into different groups. Section A compares outcomes for vehicles assigned placebo stickers and any treatment sticker with those for the control group (the constant is not included in the table). The point estimate of the placebo effect on annual claims is negative, whereas the effect on maximum speed is positive, but neither effect is significantly different from zero. However, vehicles with any nonplacebo sticker had statistically significantly fewer claims. Although we do not detect a placebo effect directly, we cannot reject equality of the treatment and placebo effects. Although we believe our evidence (on which we elaborate below) supports the claim that explicit calls for action are important motivators of behavior change, even the low dosage of the placebo may have been partially effective.

Under the parallel trends assumption, the counterfactual annualized rate of claims expected in the post period for vehicles assigned to any nonplacebo treatment was 6.86%, so the coefficient of −0.017 represents a reduction in claims of 25%.

Using GPS devices, we automatically recorded information on speeds on 4,405 separate trips undertaken by our field staff. Maximum speeds in nonplacebo treatment vehicles were on average about 1 km/h less than those of the control group in the post period, although this difference is not statistically significant. A broader measure of vehicle speeds is the average moving speed shown in the third column. Here we estimate a significant difference: vehicles with stickers are about 1 km/h slower than control vehicles.

Section B also estimates the impact of stickers motivating individual vs. collective action by comparing outcomes for vehicles in these subgroups with those in the placebo and control groups combined. The effects on claims are both negative, although statistically significant only for the collective action subgroup, for which the point estimate is about twice as large. The coefficient of −0.22 represents a 32% reduction in claims against the estimated counterfactual. We argue that the difference in focus, between individual and collective action, is unlikely to affect drivers directly, so that the differences in outcomes between the two groups suggests differences in, and indeed the presence of, actual passenger responses or the threat thereof. On the other hand, maximum speeds exhibit similar responses to the two treatments, being lower in both subgroups than in the control/placebo comparison group, although the effect is significant at the 10% level for the individual action subgroup. Average

Table 4. **Impacts of sticker assignments**

Experimental arms	Claims†	Maximum speed (km/h)	Average speed (km/h)	Reckless event	Passenger voice
A. Main effects					
Placebo	−0.014	0.908	0.146	0.027	0.049
	(0.010)	(0.745)	(0.792)	(0.033)	(0.035)
Any treatment sticker	−0.017**	−0.979	−1.019**	0.020	0.005
	(0.008*	(0.606)	(0.506)	(0.024)	(0.027)
Treatment = placebo, *P* value	0.29	0.00	0.13	0.80	0.11
B. Individual vs. collective action					
Individual	−0.012	−1.257*	−0.853	0.024	−0.003
	(0.008)	(0.649)	(0.629)	(0.026)	(0.029)
Collective action	−0.022***	−0.687	−1.194**	0.017	0.013
	(0.008)	(0.637)	(0.484)	(0.027)	(0.030)
C. Message types					
Text message only	0.001	−0.692	−1.404***	0.022	−0.012
	(0.010)	(0.707)	(0.532)	(0.032)	(0.033)
Supportive: Voice	−0.024**	−0.039	−0.984*	0.036	0.018
	(0.010)	(0.685)	(0.509)	(0.030)	(0.033)
Event aversion: Crash	−0.028***	−2.039***	−0.966	0.012	0.001
	(0.010)	(0.717)	(0.811)	(0.030)	(0.034)
Consequence aversion: Injury	−0.016*	−1.210*	−0.755	0.010	0.010
	(0.010)	(0.734)	(0.824)	(0.031)	(0.034)
Observations	23,474	4,405	4,405	1,471	1,471

Clustered SEs in parentheses. Significance at *10%, **5%, and ***1% levels.

†In the claims regression, entries are coefficients on the reported independent variable interacted with postrecruitment in a difference in difference specification. In the speed and survey regressions, entries reflect simple differences.

moving speeds are also both lower than the control group, but this time the difference is significant for the collective action subgroup.

Finally, section C reports estimated effects for each message type (combining both individual and collective action subgroups within each arm) compared with the control and placebo groups. Text-only stickers appear to have no impact whatsoever on claims, although they are associated with a reduction in average moving speed of 1.4 km/h. All other message types reduce claims by statistically significant margins and are associated with lower maximum and average moving speeds. Effect sizes range from a 26% reduction in claims (for the consequence aversion treatment) to a fall of 34% for each of the other two treatments (supportive and event aversion). Across these groups, maximum speeds are 0–2 km/h lower, and average moving speeds are 0.8–1.0 km/h lower.

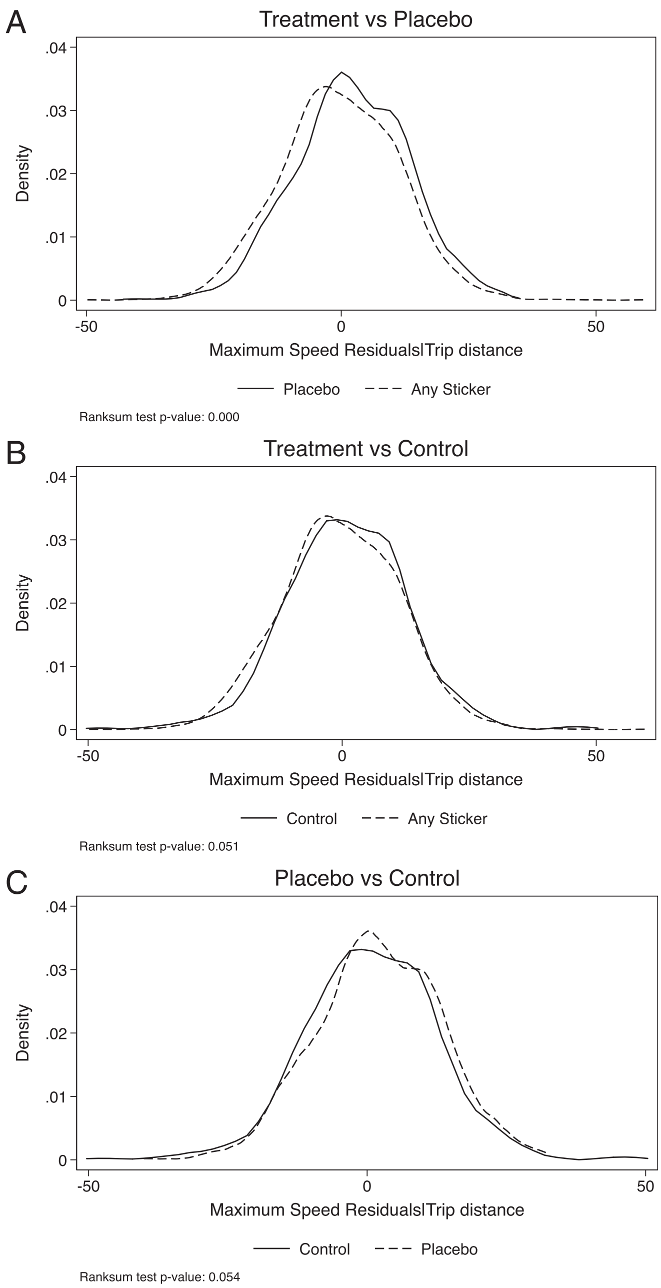

Fig. 3. (*A*) Kernel estimators of conditional maximum speed densities: treatment vs. placebo. (*B*) Kernel estimators of conditional maximum speed densities: treatment vs. control. (*C*) Kernel estimators of conditional maximum speed densities: placebo vs. control.

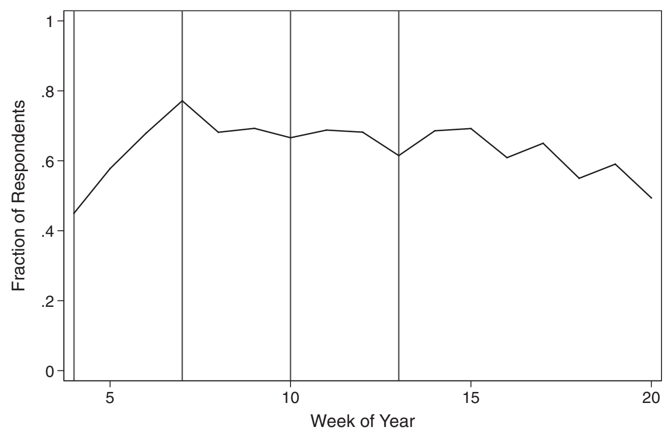

Fig. 4. Exposure to the radio campaign.

In Table 4, we are unable to detect differences in reports of reckless driving (fourth column) or passenger complaints (fifth column) across treatment groups. This lack of directly observed changes in behavior could reflect the more subjective nature of the data collection process, as well as the nature of the impacts of the stickers, on which we comment further below. We were only able to match valid qualitative data in 1,471 of the trips taken by our enumerators, who manually recorded information on the types of driving events as well as associated passenger and driver reactions.

From Table 2, on 15% of trips some kind of reckless driving was reported, whereas on 17% passengers were observed to speak to the driver. Across all groups, a little over half of all passenger complaints were due to trip interruptions such as breakdowns, but 30% reflected reckless driving. Of the latter, in two-fifths the driver was speeding, in two-fifths he was overtaking dangerously or swerving, and in one-fifth he was driving distractedly. Women and men were equally likely to speak up, and in two thirds of the events, both did so. Nine percent of those who voiced complaints were considered by our enumerators to be young, nearly half were middle aged, and the rest were elderly.[††] Overall, speaking up against bad driving does not appear to be unusual.

Our regressions reflect average impacts across all vehicles by assignment. Although claims are sufficiently rare (statistically) and necessarily discrete, making it difficult to discern distributional impacts, such impacts are more easily observed in our maximum speed data. Fig. 3 *A–C* plots pairwise comparisons of the distributions of maximum vehicle speeds for all recruited vehicles in our sample, by assignment to control, placebo, and any treatment.

Comparing the treatment and placebo distributions (Fig. 3*A*), treatment group vehicles exhibit a speed distribution that is displaced to the left, and we reject equality of these distributions at the 1% level ($P = 0.000$). This difference appears to be due to two effects: first, the distribution of treatment group vehicle speeds is shifted to the left of that of the control group—we can reject equality of these distributions (Fig. 3*B*) at the 10% level ($P = 0.051$); and second, the distribution of placebo vehicle speeds is shifted to the right of the control group—we reject equality of these distributions (Fig. 3*C*) again at the 10% level ($P = 0.054$).

[††]We did not ask the passengers their ages, but instead our enumerators were instructed to record the passengers as being young, middle-aged, or elderly.

The fact that maximum speeds in the placebo group appear to be higher than those of the control warrants some discussion. One potential explanation is that the "Travel Well" stickers might have induced slightly more aggressive driving behavior on dimensions that did not necessarily increase accidents, such as driving a little faster on the open road. Such behavior would be consistent with the higher speeds shown in Fig. 3*C*, whereas not having a demonstrable effect on accidents.

Results of the Radio Campaign. From January to April 2012, a series of five 1-wk campaigns were aired on two regional radio stations with coverage in the country's Mt. Kenya region and Nairobi. Thirty-second radio spots promoting the Zusha! message were aired in the morning and evening peak hours from Monday to Saturday and were then withdrawn for a period of 4 wk. This timing ensured we avoided any monthly seasonality effects.

We surveyed passengers of vehicles over the 4-mo period during which the radio campaign operated, in a range of cities and towns, to ascertain exposure levels to the treatment. Fig. 4 suggests high levels of awareness on the part of passengers early on in the campaign but that the salience of the ads might have fallen over time.[‡‡]

Fig. 5 shows the approximate geographic coverage of the radio stations that carried the road safety messages. The most relevant aspect of the figure is that Mombasa (on the coast) is far removed from the circled areas reached by the radio campaign.

Fig. 6 includes four panels. In each of the first three panels, we illustrate the difference between the weekly coefficients δ_{jt} of Eq. **2** in a given exposed region (Nairobi, Nakuru, and Mt. Kenya) and those estimated for Mombasa, along with the 95% confidence interval (CI). The fourth panel, for comparison, reports the point estimates of the three region weekly differentials (relative to Mombasa) in one graph and no confidence intervals. The vertical lines in each graph are placed at the weeks during which the radio campaign was aired.

There is no evidence that accident rates in any of the exposed areas consistently exhibit reductions in comparison with those in Mombasa in or soon after weeks during which the radio spots played.[§§]

What Can We Say About Mechanisms? Although we detect reduced accidents and slower speeds among vehicles assigned to treatment, we do not observe more passenger complaints in them. Of course, this does not necessarily imply that the stickers do not encourage passengers to voice concerns about dangerous driving, because both complaints and bad driving are likely to be determined simultaneously. Such a conclusion could only be drawn if we were able to exogenously randomize reckless driving itself across vehicles, orthogonal to sticker treatment assignment. However, we argue that our data are more consistent with an increase in passenger empowerment than with a direct effect on the driver.

First, the data do indicate a certain level of passenger complaints among the control group. These complaints could conceivably be random, unrelated to driver behavior, but this seems unlikely. Under this reasonable assumption, if the stickers work through a direct effect on the driver, then we should observe lower levels of complaints in the treatment groups. On the other hand, if the stickers work by inducing passengers to complain more, conditional on reckless driving, then the effect on equilibrium complaints is ambiguous: they should increase holding reckless driving constant, but fall as such recklessness is reduced (which it appears to be, from the claims and speed data). Power and data quality issues aside, our results favor a mechanism consistent with passenger empowerment ahead of a direct driver effect, although we are unable to draw definitive conclusions.

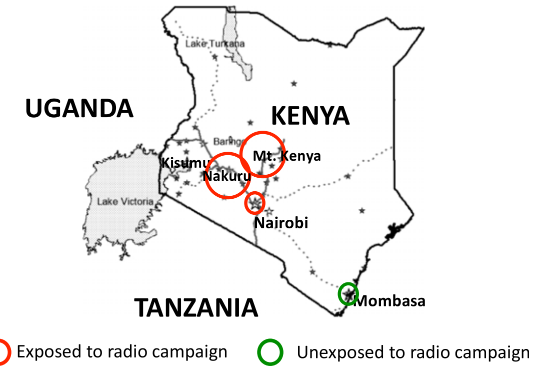

Fig. 5. Approximate areas of coverage of radio stations that carried the road safety campaign.

Financial and Social Returns to the Intervention

Among the roughly 8,000 vehicles in the treatment groups, about 140 accidents were avoided over the course of a year. Using information on the cost of each claim, we are able to calculate the implied financial rate of return to the insurance company from the intervention. The primary (nonresearch) costs of administering the intervention included the costs of printing and distributing the stickers and the costs of the lottery. On the basis of these data, we estimate a financial rate of return between 57% and 255%.[¶¶,##]

Associated with the accidents that were avoided, we estimate 55 lives were saved. From our survey, we have some information on the age distribution of matatu passengers, although we lack such information on others who might have been injured or killed in accidents, such as pedestrians. However, we conservatively estimate the cost-effectiveness of our intervention as being between $13 and $60 per DALY saved. This value is derived from our passenger surveys, which indicate an average age of about 30 y, and an assumption that life expectancy at 30 in this population is about 60 y. We further assume that disability adjustments mean that each injury avoided saves the equivalent of between 5 and 15 y of healthy life and that the ratio of injuries to deaths is between 2 and 10–1.[***]

[‡‡]The high levels were reported in response to questions about ever having heard the radio spot. When asked if they had heard it "yesterday," respondents reported lower rates of exposure, consistent with the interrupted deployment of the radio campaign.

[§§]Restricting this analysis to those vehicles that were observed to be insured with our partner insurance company in at least 8 of the 16 mo between March 2011 and July 2012 yields similar results.

[¶¶]The median claim cost was about 90,000 Kenyan shillings, or about $1,125 at the prevailing exchange rate at the time, whereas the mean cost was 202,000 shillings, or $2,525. The costs of printing the stickers and running the lottery amounted to about $100,000. The financial rate of return using the median cost was 57%, and using the mean cost was 255%.

[##]To the extent that both the intervention and its effect depreciated over time, this figure would also represent the return to a sustained intervention in which stickers were redesigned and replenished, although of course it is difficult to predict the behavioral response of passengers over the longer term. Either they could become immune to the intervention, in which case it would lose its effect, or the behavior it motivates could constitute a new social norm, in which case there would be no need for continued exposure and the return (both financial and social) to the initial investment would be enormous.

[***]Even these numbers are conservative. The WHO considers the ratio of injuries to deaths to be between 20 and 50 to 1.3 (see above). Our data suggest a ratio of 10 to 1, although a large fraction of injuries are relatively minor soft tissue injuries.

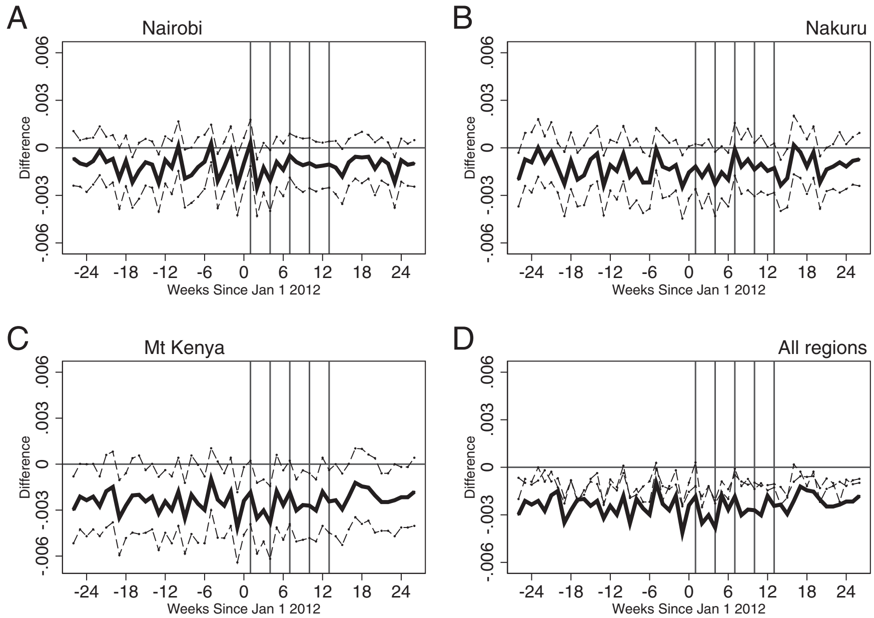

Fig. 6. Radio campaign analysis. (*A*) Differential weekly claims rates in Nairobi vs. Mombasa (with 95% CI). (*B*) Differential weekly claims rates in Nakuru vs. Mombasa (with 95% CI). (*C*) Differential weekly claims rates in Mt. Kenya vs. Mombasa (with 95% CI). (*D*) Point estimates of differential weekly claims rates in Nairobi, Nakuru, and Mt. Kenya vs. Mombasa.

Although these estimates depend on a number of underlying assumptions, they suggest the intervention is highly cost-effective in saving lives and comparable to some of the "best buys" in public health. The more impactful stickers, which were associated with reductions in accident rates of 34%, were correspondingly more cost-effective: for these stickers, the cost per DALY saved was between $10 and $45.[†††]

Conclusions

Our results suggest that salient and actionable information delivered in a timely manner can be effective in improving the safety of public bus travel in Kenya and by extension in other contexts. Overall, reported accidents fell by a quarter in vehicles assigned to the treatment groups, and messages that promoted collective action against bad driving reduced claims by one-third, as did messages with particularly evocative images. Text-only messages were particularly ineffective, possibly due to limited literacy, but we expect more for psychological and emotional reasons, especially considering the adult literacy rate in Kenya is 72%[‡‡‡] and likely higher in the areas in which the study vehicles operated. Although strikingly large, the effect sizes were somewhat smaller than the even larger 50% reduction observed in our 2011 study. That study, however, was conducted with long distance vehicles only, whereas the current evaluation included both long distance and intracity buses. We hypothesize that city accidents could be less responsive to the sticker intervention, because speeds are often much lower, as are the costs of accidents perceived by passengers, and hence their proclivity to speak up even when nudged to do so.

More ubiquitous information, as delivered over a radio campaign, had no discernible effect in reducing claims, although the impacts of a longer-term and more intense campaign could be larger.

Although we cannot detect differences in actual passenger behavior between treatment and control groups, the striking effectiveness of the collective action intervention, which we hypothesize is unlikely to differentially affect the driver's perceptions or actions, suggests that consumer empowerment lies at the heart of the intervention's impact. Also, we argue that the lack of correlation between treatment assignment and observed passenger behavior is inconsistent with a direct driver mechanism. Buses assigned to all treatment groups are observed to reach maximum speeds that are on average about 1–2 km/h slower.

The impacts of our intervention on both maximum and average moving speeds and claims allow us to robustly calibrate the causal relationship between speed reduction and accidents, because the variation is exogenously driven by treatment assignment. The results support and provide perhaps the first, to our knowledge, rigorous confirmation in a sub-Saharan African context of, the WHO's (1) claim that lowering speeds by 5 km/h can reduce accidents by 25%. Indeed, our results suggest either that smaller reductions in speed are sufficient to attain such safety outcomes and that lowering speeds by 5 km/h could have even larger impacts than anticipated or that our intervention both lowered speed and reduced other forms of recklessness, such as dangerous overtaking or distracted driving, which both combined to reduce the rate of accidents substantially.

ACKNOWLEDGMENTS. We thank Woubie Alemayehu, Jon Bernt, Anne Davis, James O'Sullivan, Sree Papineni, and Ali Sharman for dedicated field management and research assistance. We acknowledge support from the United States Agency for International Development, the Safaricom Foundation, the National Bureau of Economic Research Africa Project, and Georgetown University.

[†††]An alternative measure of the efficacy of the intervention measured the lives saved against the time lost due to lower speeds. Based on the average trip length (about 60 km) and a reduction in average speed of 1 km/h, each person loses 1 min per trip. Assuming there are 3 trips per day and 14 passengers (the legal limit), this amounts to 42 min per vehicle per day. Over 365 d, the equivalent of 10.6 d is thus lost per vehicle in a year. For the 8,000 vehicles in our treatment groups, a total of 233 y of time lost. If a death is equivalent to the loss of 30 y, lost time per year amounts to the same as 7.8 lives: i.e., slowing down costs 7.8 lives of time, but in return 55 lives are saved.

[‡‡‡]Unicef, www.unicef.org/infobycountry/kenya_statistics.html.

1. World Health Organization (2013) *Global Status Report on Road Safety 2013* (WHO, Geneva).
2. Habyarimana J, Jack W (2011) Heckle and chide: Results from a randomized road safety intervention in Kenya. *J Public Econ* 95(2011):1438–1446.
3. Guria J, Leung J (2004) An evaluation of a supplementary road safety package. *Accid Anal Prev* 36(5):893–904.
4. World Health Organization (2009) *World Report on Road Traffic Injury Prevention* (WHO, Geneva).
5. Afukaar FK, Antwi P, Ofosu-Amaah S (2003) Pattern of road traffic injuries in Ghana: implications for control. *Inj Control Saf Promot* 10(1-2):69–76.
6. Bishai D, Asiimwe B, Abbas S, Hyder AA, Bazeyo W (2008) Cost-effectiveness of traffic enforcement: case study from Uganda. *Inj Prev* 14(4):223–227.
7. Mathers CD, Loncar D (2006) Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med* 3(11):e442.
8. Odero W, Khayesi M, Heda PM (2003) Road traffic injuries in Kenya: Magnitude, causes and status of intervention. *Inj Control Saf Promot* 10(1-2):53–61.