Experimentally Estimating the Value of a Statistical Life:

Pre-analysis Plan for wave 2[*]

Grady Killeen

University of California, Berkeley

February 6, 2023

Abstract

Summary: This document outlines outcomes and regressions for estimating (1) the reduced-form effect of safety information on demand for motorcycle helmets in Nairobi, Kenya, and (2) the value of a statistical life, estimated using randomized variation induced by this experiment. The study is a lab in the field experiment in which passengers of motorcycle taxis in Nairobi will be presented with a randomized debiasing intervention containing information about the empirical risk of a fatal accident and the efficacy of helmets at preventing death. The study consists of a pure control group that will not be asked about perceived risk, a control group that will be asked about perceived mortality risk but not presented with information, a treatment group in which respondents are presented with empirical mortality data and an academic study estimating that helmets reduce one's probability of dying by 42%, and a treatment group in which respondents are presented with empirical mortality data and an academic study estimating that helmets reduce one's probability of dying by about 70%. All respondents will then participate in a Becker et al. (1964) willingness to pay exercise.

Appendix A: Survey instrument in English.

---

1 Wave 2 amendments

This document presents an updated pre-analysis plan filed between waves 1 and 2 of data collection. This section (including subsections) details deviations from the pre-analysis plan filed before wave 1 informed by the initial round of data collection. Following this section, the original pre-analysis plan is included in full for reference.

1.1 Wave 2 sample and timeline

The target population and recruiting strategy are unchanged for wave two compared to wave one of the study. The sample remains restricted to adult motorcycle taxi passengers that do not report owning a helmet, and respondents are recruited from motorcycle taxi stands throughout the Nairobi metropolitan area.

Wave 1 included 921 completed surveys. 93 surveys (9%) were pure control, 258 were assigned to control (28%), 294 were assigned to low treatment (32%), and 286 were assigned to high treatment (31%). Details about the treatment groups are included in the original PAP below.

We plan to eliminate the pure control group for the second wave and are targeting around 760 additional surveys. Respondents will be assigned to one of the three remaining arms with equal probability. Hence, modifications from the original PAP related to estimation of the value of a statistical life (VSL), which does not use pure control observations, were made after just over half of the data we anticipate using for analysis has been viewed by the research team.

The first wave of the survey was implemented from October - December, 2022. We plan to implement the second wave in February and March, 2023.

1.2 Excluded data from wave 1

During one day of data collection, surveyors reported that motorcycle taxi drivers, who are excluded from the sample, were reporting to be motorcycle taxi passengers to participate in the survey and obtain free helmets. This occurred on the second day of data collection when some surveyors returned to locations visited during the pilot and motorcycle taxi operators recognized the survey team and were familiar with the sample restrictions. Some motorcycle taxi operators parked their bikes and had colleagues drive them around the block so that they appeared to be passengers. After completing surveys, the surveyors recognized the drivers later in the day when they were transporting passengers.

Surveyors informed the field manager and principal investigator about the problem before data was viewed. Inspection confirmed extreme outliers, particularly with respect to helmet valuations, that are consistent with heterogeneity in ridership or false responses. Hence, we plan to exclude all 35 observations from the day of data collection from our primary estimates since they may be contaminated. We will report results with the observations in an appendix.

Following the issues of contamination, we instructed surveyors to focus on recruiting passengers approaching on foot, perform surveys further from motorcycle taxi stands, and carefully observe who the operators were when arriving to exclude them from surveys. There were no similar issues reported on later days.

1.3 VSL instrument changes

Prior to the first wave of data collection, we pre-specified two 2 stage least squares regressions for estimating the value of a statistical life. The first is

$$v_i = \zeta_h + VSL\Delta r_i + X_i'\gamma_0 + \gamma_1 n_i + \epsilon_i$$

$$\Delta r_i = Z_i'\pi + \nu_i$$

where Δr_i denotes the reduction in mortality risk that a respondent perceives over the 5 year lifespan of a helmet, v_i is a respondent's helmet valuation measured via the BDM exercise and X_i represents a vector of controls. n_i is the number of motorcycle taxi trips that a passenger takes in a typical week. The vector of excluded instruments is given by $Z_i = (1, n_i, T_i', n_i \cdot T_i', p_i, X_i')$ where T_i is a vector of indicators for treatment status.

The reason for including n_i and the interaction of n_i with treatment assignment was to improve power. Intuitively, if there is large heterogeneity in the risk of riding motorcycle taxis across the population, then Δr_i will vary both due to heterogeneity in risk and beliefs about the efficacy of helmets. Hence, if heterogeneity in perceived motorcycle accident risk is large it will be harder to detect an effect of the treatment on Δr_i.

We initially aimed to use n_i as a control under the basis that, holding all else equal, someone that rides a motorcycle more often will have a higher risk of suffering a fatal accident. In practice, however, we found that pre-treatment beliefs about fatal accident risk were poorly correlated with ridership. The R^2 from a regression of baseline beliefs on ridership is below 0.0001 in wave 1 and the p-value is 0.969.

Hence, we plan to estimate a similar model replacing n_i with the respondents' baseline beliefs

3

about their mortality risk over the following 5 years. This is measured consistently across those in the treatment and control groups before treatment assignment is revealed to the surveyor, so it forms a valid instrument. The intuition for this model is the same as that with n_i and $n_i \cdot T_i$, but it is more efficient since it better absorbs heterogeneity in baseline beliefs.

The second set of instruments we pre-specified in the initial PAP uses an intercept and T_i only. This model is unchanged.

1.4 Control variable changes

The initial PAP specified that a rich set of demographics plus baseline ridership would be considered as potential controls, with LASSO used to select. Furthermore, we specified that all regressions would control for baseline ridership.

Given the switch to the use of baseline fatal accident beliefs for the number of trips, we plan to instead include baseline beliefs rather than ridership volume in all estimates. Furthermore, due to an error in the initial PAP we failed to include the full set of potential controls for VSL estimates. In addition to demographic variables, we include questions about the respondent's baseline beliefs about motorcycle taxi risks, the information sources they used to form these beliefs (e.g. own experiences, news sources, government data), their confidence in their initial beliefs, the types of trips that they take (e.g. commuting or leisure), and the reason they take motorcycle taxis rather than alternative modes of transportation (e.g. congestion or cost). We plan to include these covariates, which are measured before treatment status is revealed, in the set of potential controls when estimating VSL.

Moreover, we may include enumerator fixed effects. Different surveyors worked in different areas of Nairobi, leading to heterogeneity that is associated with surveyor. Furthermore, we found

4

in wave 1 that some surveyors are better than others at helping respondents think about and con-textualizing risks, leading to further variation explained by the surveyor. Hence, surveyor fixed effects can help explain variation in helmet valuation and improve power.

The remainder of this document contains the originally filed PAP.

1.5 Survey changes

We do not plan to make any survey changes that affect outcomes of interest in the study for wave 2. We do plan to add common "other" responses to answer options. In addition, we added a module exploring the role of search costs in low helmet use at the end of the survey. This is primarily exploratory.

2 Introduction

This document outlines the analysis plan for a lab in the field experiment that aims to experimentally estimate the value of a statistical life in Nairobi, Kenya. The study focuses on passengers of motorcycle taxis, called "boda bodas" or "bodas." The core component of this intervention is to present passengers that are not wearing helmets with information about the mortality risk of bodas in Kenya and the efficacy of helmets. I then plan to conduct a willingness to accept exercise to measure demand for helmets using a Becker et al. (1964) mechanism (hereafter referred to as BDM). This pre-analysis plan is being filed after a short pilot aimed at fixing issues with the survey instrument. None of the data that will be used in analysis has been seen by anyone on the research team at the time this is filed.

This project has two primary aims. First, from a public policy perspective the study aims to examine the role that prices and information have in uptake of motorcycle helmets. Bachani et al. (2017) found in an observational study that less than 3% of boda passengers wear helmets, despite the fact that traffic accidents are the leading cause of death for individuals 18-25, and motorcycles are particularly risky. Meanwhile, academic studies suggest that helmets are effective at reducing mortality risk. Liu et al. (2008) conduct a meta-analysis, primarily from developed

contexts, producing a point estimate of 42% efficacy. Ouellet and Kasantikul (2006) estimate over 70% efficacy in Thailand, a setting that may be more similar to Kenya. Hence, helmets are among the most significant investments that frequent boda passengers can make to reduce their mortality risk.

This project aims to build on a recent economics literature such as Habyarimana and Jack (2011) demonstrating that information treatments can yield behavioral change that is effective at reducing traffic deaths. Specifically, participants will be assigned to a pure control group that is not asked any questions about mortality risk or helmet effectiveness, a control group that is asked a series of questions to measure mortality risk beliefs, a treatment group presented data from Liu et al. (2008) showing 42% effectiveness, and a treatment group given the finding from Ouellet and Kasantikul (2006) that helmets reduce the odds of a fatal accident by 70%.

All treatment groups will then participate in a BDM willingness to accept exercise. Respondents will first be asked to state the smallest cash payment, in Kenyan shillings, they would prefer to a free helmet. We will then select a payment value between 5 and 600 shillings with uniform probability. Respondents will receive the payment if the draw is larger than their stated valuation and otherwise receive the free helmet. The study uses a willingness to accept exercise rather than a willingness to pay mechanism to prevent liquidity constraints from binding.

The helmets we are offering sell at a wholesale price of Ksh 580 per unit from the Kenyan manufacturer Boda Plus. Based on discussion with an NGO and a helmet manufacturer, this appears to be on the upper end of helmet prices due in part to an abundance of low quality and counterfeit products. Boda Plus is a subsidiary of Car and General, one of the largest motorcycle sellers in East Africa, and it produces helmets that adhere to the Kenyan Bureau of Standards' safety requirements which map closely to standards created by the United Nations. In addition, an unaffiliated

NGO that advocates for helmet safety in Kenya indicated that the helmets are high quality.

This design allows us to estimate the reduced form effect of safety information on helmet valuations. Moreover, the BDM exercise allows one to determine what share of respondents would purchase a helmet at different prices, providing insight about the efficacy of price subsidies and complementary effects with information campaigns that are relevant to policy makers. Details about the regressions that I plan to estimate are presented in section 5.

The second aim of this project is to use the randomized variation induced by the intervention to estimate the value of a statistical life (VSL), that is willingness to pay to reduce mortality risk. This study aims to make several contributions to the literature. First, VSL estimates generally depend on the assumption that individuals have unbiased beliefs about the mortality risk of the decision that they face. This assumption is strong since mortality risks are not easily observable. This experiment allows us to directly test whether mortality beliefs are systematically biased since both reported beliefs and valuations should be orthogonal to the information treatment under the null hypothesis that beliefs are correct. Moreover, we can estimate demand models separately on the information treatment and control groups using endogenous variation in mortality risk reduction from a helmet associated with how often one rides a boda. This mirrors the type of analysis often used to estimate VSL, and allows us to see whether biased beliefs significantly change estimates.

In addition, the experiment allows for an estimate of the VSL of urban Kenyans using experimental variation. The information treatment will create exogenous variation in participants' perceived reduction in mortality risk from purchasing a helmet. Combined with their valuations measured using the BDM exercise, we may estimate a demand system for helmets that identifies VSL. The average value of a statistical life is simply the coefficient on the mortality risk reduction in an instrumental variable regression of valuation on risk reduction, instrumenting for the risk re-

duction using information treatment assignment. A simple model illustrating how the experiment identifies VSL is presented in section 4 and estimation details are in section 6.

3   Sample, Study Design, and Data

This study will consist of a single survey, included in the appendix for reference, that will take between 15 and 45 minutes. Respondents will be recruited at boda stands in Nairobi, locations where individuals go to obtain a motorcycle taxi ride. Surveyors will interview respondents throughout the day, although traffic is generally higher during morning and evening commutes. We expect that many potential respondents will be time constrained and thus aim to minimize the duration of the survey. Surveyors will visit multiple boda stands throughout the city in order to reach a broader sample of passengers.

The sample of boda passengers is likely selected. For instance, those with less risk aversion or a higher value of time may be more likely to take motorcycle taxis. This may limit the external validity of the study, particularly VSL estimates. However, boda use is increasingly common across demographics due to congestion, so this design is likely to reach a broader segment of the population than many studies of VSL. I also plan to collect detailed demographic data from respondents. Demographic information is obtained prior to any randomized components of the experiment. I may examine corrections, such as control function approaches or weighting, to estimate VSL values for the broader population of Nairobi or of Kenya, rather than just the sample of motorcycle taxi passengers.

Data collection will consist of approximately seven weeks of data collection. This follows a one week pilot aimed at refining questions aimed at measuring beliefs about the mortality risks of motorcycle taxis.

This wave of data collection is constrained by the budget for this round of data collection rather than a target sample size. Conducting ex-ante power calculations requires strong assumptions because the sensitivity of consumers to helmet prices, baseline beliefs about helmet effectiveness, and baseline beliefs about mortality risk are all unknown. Thus one needs to make arbitrary assumptions about effect sizes to calculate statistical power. In addition, a core concern for estimating VSL is instrument strength rather than simply detecting an effect size. Given the barriers to producing credible power calculations and the fact that this study received a pilot grant for this wave of data collection, I plan to collect the largest sample possible under the current funding. This will likely produce a sample size between 800 and 1,200. I then plan to estimate the regressions reported in this pre-analysis plan and perform power calculations via simulation. I will then apply for funding to collect an additional sample in a second wave if necessary and pool the sample from the two rounds of data collection, including wave fixed effects in all analysis.

Data will be collected using SurveyCTO. Randomization will be conducted in SurveyCTO using the *random()* function. This function uses the Java randomization algorithm to take a pseudo-random draw from a standard uniform distribution. This study does not stratify randomization since the sample is not known en-ante. Hence, randomization must be conducted in the survey. Independent random draws are used to determine which information treatment group the respondent is assigned to and which price offering the respondent receives in the BDM demand exercise. Since a primary focus of this study is estimating the value of a statistical life, which requires data on individuals mortality beliefs, respondents are assigned to the pure control with a lower probability. We plan to offer respondents a cash payment between Kenyan shillings (Ksh) 5 and 600 with uniform probability. Respondents are informed about the range of possible cash payments during the consent process. However, we do not include this range when introducing the BDM

mechanism to avoid confusion about the game since the range does not affect optimal strategies and to avoid creating anchor points.

Table 1: Information treatment probabilities

Treatment group	Assignment probability
Pure control	10%
Control	30%
Treatment 1: Liu et al. (2008)	30%
Treatment 2: Ouellet and Kasantikul (2006)	30%

The first information treatment group is a pure control. These respondents will be asked a series of demographic questions and basic information about boda ridership, then proceed to the willingness to accept exercise. Those assigned to the pure control will not be asked any questions about their perceived likelihood of dying in a motorcycle accident, or other questions relating to motorcycle safety. The aim of the pure control group is to provide an estimate of baseline demand for motorcycle helmets among consumers that are not asked to think carefully about safety before being offered a helmet.

The control group will be asked detailed questions about their perceptions about boda safety. We currently do not plan to offer any information about mortality risk or helmet effectiveness to those in the control group. However, we may present those in the control group with data about the respondent's empirical accident risk depending on responses during the first several weeks of piloting. If respondents have extremely diffuse priors about empirical mortality risk, then noise in the measurement of perceptions about accident risks may dramatically reduce the power of this study. In this case, we may present the control group with empirical accident risk data so that the primary variation generated by the study will enter through beliefs about helmet effectiveness which are well-studied, more heavily publicized, and easier to measure. If we present those in the

11

control group with empirical risk estimates at some point in the study, we will include fixed effects to capture the change in treatment.

Empirical risks are calculated using the 2021 mortality risk per motorcycle trip of motorcycle drivers, since we know that one driver is present during each trip but were not able to obtain high quality estimates about the frequency of trips that involve a passenger, calculated using data from news sources and the Kenyan National Transport and Safety Authority (NTSA). We then use the per-trip risk of an average Kenyan and the respondent's ridership volume to estimate their mortality risk over a 5-year span, the recommended lifespan of the helmet.

Both treatment groups will receive information about helmet effectiveness. The first treatment group will be presented with the results of Liu et al. (2008) which conducts a meta-analysis of studies on helmet effectiveness, predominately from developed contexts, and estimates that helmets reduce mortality risk by 42%. Those in the second treatment group will be presented with the finding from Ouellet and Kasantikul (2006) that helmets in Thailand reduced mortality risk by roughly 70% when properly worn.

We plan to measure mortality beliefs by first asking for per-trip estimates of risk for a standard Kenyan. We will then ask for the risk per 1 year and per 5 years of a Kenyan that utilizes bodas as frequently as the respondent. We will next ask what the respondent believes their own risk is over the next 5 years. This ordering aims to help respondents think critically about the risk, and asking about an average Kenyan and then the respondent's own risk may help the respondent think about their own risk. Respondent's beliefs about their own mortality risk is measured using a two-step approach in which we ask respondents to first select from a list of risk ranges and then provide a more granular estimate within the selected range.

We will then present the empirical 5-year mortality risk, if applicable, before finally presenting

12

each of the 5 year estimates to the respondent and asking them to produce a final estimate. The aim is to help respondents refine an estimate by approaching the question in multiple ways, and then allowing them to select the most credible estimate. However, the volume of questions is time consuming and respondents may have an aversion to being asked a large volume of sensitive questions, so we may reduce the number of questions based on piloting.

Respondents in the treatment group will then be presented with study results about helmet efficacy, then those in the control and treatment groups will be asked about their own beliefs about helmet effectiveness. We will then use this value and the estimated 5-year mortality risk without a helmet to present the respondents with the risk reduction offered over the lifespan of the product. This makes mortality risk salient when the respondent is considering the value of a helmet relative to cash, making VSL estimates more credible. It also helps reduce the cognitive burden of calculating the mortality reduction from a helmet. This is likely to reduce noise in outcomes.

The final risk reduction stated to the respondent will generally be used in regressions. However, we may also consider other measures of mortality risk collected in the survey, such as the risk an average Kenyan faces if respondents are uncomfortable thinking about their own risk.

4 Identification of the Value of a Statistical Life

This section presents a simple model illustrating how the value of a statistical life (VSL) is identified from this experiment.

Suppose that a consumer has a prior about the probability of dying with a helmet in an accident that would be fatal without a helmet given by

$$Pr(D|H; \mathcal{I}_0) \sim Beta(\alpha_{0H}, \beta_{0H})$$

13

where $\mathcal{I}_0$ denotes the individual's baseline information set. Then

$$H_0 \equiv \mathbb{E}[Pr(D|H;\mathcal{I}_0)] = \frac{\alpha_{0H}}{\alpha_{0H} + \beta_{0H}}$$

Now suppose that the consumer is presented with information that the estimated efficacy of helmets is $\theta_H \sim Beta(\alpha_{EH}, \beta EH)$. Then their posterior beliefs about the efficacy of helmets are given by

$$Pr(D|H;\mathcal{I}_1) \sim Beta(\alpha_{0H} + \alpha_{EH}, \beta_{0H} + \beta EH)$$

and the expected value is

$$H_1 \equiv \mathbb{E}[Pr(D|H;\mathcal{I}_1)] = \frac{\alpha_{0H} + \alpha_{EH}}{\alpha_{0H} + \alpha_{EH} + \beta_{0H} + \beta EH}$$

If $\frac{\alpha_{0H}}{\alpha_{0H}+\beta_{0H}} \neq \frac{\alpha_{EH}}{\alpha_{EH}+\beta_{EH}}$, the consumer initially has biased beliefs and their posterior mean will differ from their prior mean. The degree to which their posterior will update depends on the magnitude of bias in initial beliefs, how diffuse their prior is, and how diffuse the signal is. By measuring beliefs before receiving the signal and the agent's confidence in their prior, we may thus determine the credibility with which respondents perceive debiasing information based on the extent to which their posteriors update. Furthermore, if consumers are given two different signals about the efficacy of helmets, in this case from the two different studies used in treatment 1 vs treatment 2, then their posterior means will differ so long as their prior is non-singular.

Similarly, suppose that the agent has a prior distribution about the probability of getting into a

fatal motorcycle accident, per trip, given by

$$Pr(A|\mathcal{I}_0) \sim Beta(\alpha_A, \beta_A)$$

Then their prior expectation of the accident risk per trip is given by

$$A_0 = \frac{\alpha_A}{\alpha_A + \beta_A}$$

and after receiving a signal about the empirical accident risk is θ_A their posterior mean will be given by

$$A_1 = p_A \cdot \left(\frac{\alpha_A}{\alpha_A + \beta_A}\right) + (1 - \rho_A) \cdot \theta_A$$

Suppose the consumer completes n boda rides over the course of the lifespan of a helmet. Then their baseline expectation of mortality risk without a helmet is given by

$$r_{in} = 1 - (1 - A_0)^{n_i}$$

Under prior beliefs with a helmet, it is given by

$$r_{ih0} = 1 - (1 - A_0 H_0)^{n_i}$$

And after updating beliefs, the perceived mortality risk with a helmet is

$$r_{ih1} = 1 - (1 - A_1 H_1)^{n_i}$$

Letting p_i denote the price of a helmet and r_{ih}, r_{in} denote the consumer's belief about mortality their risk with and without a helmet respectively, a consumer's expected utility from purchasing a helmet is

$$U_{ih} = \zeta_h + \beta(1 - r_{ih}) - \alpha p_i + \epsilon_{ih}$$

where $v = \frac{\beta}{\alpha}$ is the value of a statistical life and ϵ_{ij} represents components of utility unobserved to the econometrician. Without a helmet, the consumer's expected utility is

$$U_{in} = \zeta_n + \beta(1 - r_{in}) + \epsilon_{in}$$

Let $y_i = 1$ if the consumer purchases a helmet. Denote $\Delta r_i = r_{in} - r_{ih}$. Normalizing $\zeta_n = 0$ and assuming that $\epsilon_{ij} \sim iidEV1$, we have

$$Pr(y_i = 1|\mathcal{I}) = \frac{\exp\{\zeta_h + \beta \Delta r_i - \alpha p_i\}}{1 + \exp\{\zeta_h + \beta \Delta r_i - \alpha p_h\}}$$

To estimate VSL, researchers use data on y_i, r_{in}, r_{ih} and p_i to estimate such a demand system. A common assumption in the VSL literature is that individuals belief about mortality risk is equivalent to the empirical risk. In this context, this assumption would be that $\Delta r_i = r_i^*$ where r_i^* is the empirical risk. However, if beliefs are biased, then this approach will return a biased estimate of VSL. This follows since Δr_i is a function of the consumer's information set and $\frac{\partial Pr(y_i=1)}{\partial r_{ih}} < 0$. So if a consumer believes helmets are ineffective, estimating VSL with r_i^* would lead us to conclude that they have a low willingness to pay to reduce mortality risk.

In this context, we are able to leverage the fact that we measure precise valuations through the

BDM mechanism. Let v_i denote one's valuation, then

$$\zeta_h + \beta \Delta r_i - \alpha v_i + \epsilon_{ih} = \epsilon_{in}$$

$$\alpha v_i = \zeta_h + \beta \Delta r_i + \epsilon_{ih} - \epsilon_{in}$$

$$v_i = \zeta_h + VSL \Delta r_i + \epsilon_i$$

Hence, we may estimate regressions of valuation on Δr_i to recover the value of a statistical life.

Under the assumption that $\mathbb{E}[\Delta r_i \epsilon_i | X_i] = 0$ where X_i is a rich set of demographic control variables, we may separately estimate VSL using OLS and empirical mortality risk among subsamples that received different information treatments to test how much debiasing mortality beliefs affects VSL estimates. While these estimate rely on strong assumptions, they are common in the literature and thus offers a reasonable thought experiment. We think that these estimates of VSL will likely be biased, but the variability across treatment groups nonetheless offers insight into the extent to which biased priors could affect existing estimates.

This study further aims to estimate the value of a statistical life using exogenous variation induced by the experiment. Let T_i be a vector of binary variables equal to 1 if a respondent received a particular information treatment. By randomization, $\mathbb{E}[\epsilon_i | T_i] = 0$. Assuming that the information treatments induce a change in beliefs, $\mathbb{E}[T_i \Delta r_i] \neq 0$. Hence, we may estimate VSL using a two-stage least squares regression in which we instrument for the respondent's stated beliefs, Δr_i, using the information treatment assignment.

Since this approach uses instrumental variables, measurement error in reported mortality beliefs will not bias VSL estimates, a concern that has contributed to prior work using empirical risk

data. Intuitively, this study is leveraging the malleability of biased beliefs to generate a valid instrument to identify VSL. We aim to compare this experimental VSL estimate to those estimated from ridership volume to test how much endogeneity, both from biased mortality beliefs and omitted variable bias, affects VSL estimates in this setting. Since the non-experimental methods closely mirror those used in existing studies, we think this comparison offers insight into the credibility of existing VSL estimates, even in other contexts.

5 Public policy outcomes of interest

This section presents the reduced-form public policy outcomes that this study plans to analyze. The next section presents the structural specifications that will be considered for estimating the value of a statistical life. I plan to consider heterogeneity based on income, age, an indicator for whether the respondent has any children, performance on a digit span recall test, and life expectancy.

We do not believe that multiple testing is a significant threat in this setting since the experiment is designed to test a specific set of hypotheses and the number of outcomes is small. However, we will report Romano-Wolf corrected p-values for the outcomes in this section, which controls the family-wise error rate (FWER) using a correction based on randomization inference (Romano and Wolf, 2005). Corrected p-values will be found by first obtaining p-values on all tests using the actual data, then permuting treatment assignment and calculating adjusted p-values using the (Romano and Wolf, 2005) procedure. This allows us to correct across various tests statistics and uses a pivotal statistic.

We will not apply a multiple testing procedure to VSL estimates. This is because we are interested in estimating a specific structural parameter, not testing whether VSL is greater than 0. Furthermore, we view VSL estimation as the primary aim of this study.

18

5.1 Primary outcome: The effect of information on helmet demand

We will first estimate the reduced-form effect of the information treatments on helmet valuations.

We will first estimate intent-to-treat estimates of the form

$$v_i = \beta_0 + \sum_{j=1}^{3} \beta_j T_{ij} + X_i' \gamma + \epsilon_i$$

where v_i is helmet valuation measured via the BDM exercise, T_{ij} indicates assignment to the control group, treatment 1, or treatment 2 (measured relative to the base of pure control), and X_i is a vector of control variables. I plan to select controls using double-post LASSO (Belloni et al., 2014).

The coefficient on T_{i1}, an indicator for being in control group, tests whether respondents have a higher helmet valuation when considering mortality risk. $\beta_2 - \beta_1$ gives the effect of the low debiasing treatment on valuations, $\beta_3 - \beta_1$ yields the effect of the high debiasing treatment on helmet valuations, and $\beta_3 - \beta_2$ offers the change in valuations when respondents are given evidence that helmets are 70% effective versus 42% effective. We plan to test each hypothesis using t-tests or Wald tests. Moreover, we plan to test the joint hypothesis that information does not affect valuations using an F-test.

5.2 Primary outcome: Price semi-elasticity of demand for helmets

We aim to estimate the price semi-elasticity of demand for helmets using the regression

$$y_i = \beta_0 + \alpha \log p_i + \sum_{j=1}^{3} \beta_j T_{ij} + X_i' \beta + \epsilon_i$$

where p_i is the amount of cash offered to individual i and y_i equals one if the respondent selected a helmet at the randomly selected cash offering. α is the price semi-elasticity of demand.

We may also calculate the price elasticity of demand using valuation data. In particular, we will plot a demand curve using the valuation data and the local elasticity of demand estimated using local polynomial regression.

5.3 Primary outcome: Elasticity of willingness to pay for a helmet with respect to effectiveness

We will estimate the elasticity of demand with respect to perceived helmet effectiveness by estimating the two-stage least squares regression model

$$\log v_i = \beta_0 + \beta_r \log h_i + X_i'\beta + \epsilon_i$$

$$\log h_i = \pi_0 + \sum_{j=2}^{3} \pi_j T_{ij} + X_i'\pi_c + \nu_i$$

where h_i is the respondent's belief about how effective helmets are at preventing death, conditional on getting into an accident. The regression is estimated only using observations in the control and treatment groups since risk perceptions are not measured in the pure control group. β_r is the outcome of interest.

6   VSL outcomes of interest

This section presents details about the VSL estimates that this paper plans to produce. Let Δr_i denote the reduction in mortality risk that a respondent perceives over the 5 year lifespan of a helmet, v_i denote a respondent's helmet valuation measured via the BDM exercise and X_i denote

a vector of controls. n_i is the number of boda trips that a passenger takes in a typical week. Let $Z_i = (1, n_i, T'_i, n_i \cdot T'_i, p_i, X'_i)$ where T_i is a vector of indicators for treatment status. This will vary based on which subset of the sample is being used.

Our primary approach is to estimate the two-stage least squares regression

$$v_i = \zeta_h + VSL\Delta r_i + X'_i\gamma_0 + \gamma_1 n_i + \epsilon_i$$

$$\Delta r_i = Z'_i\pi + \nu_i$$

where VSL, the coefficient on Δr_i, is the object of interest. This estimator does not require assuming that errors are drawn from a known distribution. We include the number of trips interacted with treatment status since, controlling for the number of trips, this value is exogenous and likely to improve instrument strength. Intuitively, if someone rides motorcycle taxis more, then learning that a helmet is very effective will produce a larger percentage point change in the likelihood that a helmet will save their life than someone that infrequently rides them. If estimates are well-powered without including information on ridership volume, we may also report these specifications for robustness.

We plan to collect data on age, gender, education, income, the number of children, health, life expectancy, and performance on a digit span recall test. All VSL estimates will also control for the number of boda trips the respondent took last week since this has a significant effect on mortality risk and is necessary to include this value interacted with treatment as an instrument. I plan to select additional covariates to include in VSL estimates using double-post LASSO (Belloni et al., 2014).

I plan to consider heterogeneity with respect to age, gender, income, expected future income,

life expectancy, health, an indicator for whether the respondent has children, and performance on a digit span recall test. We plan to test each dimension of heterogeneity separately due to power constraints. However, we may consider jointly estimating heterogeneity if there is sufficient power.

We plan to estimate VSL using two samples: observations from the control arm and both treatment arms, and observations from the two treatment arms only. We cannot use pure control observations since we do not collect mortality beliefs among this sample.

We will likely winsorize mortality risk data at the 2nd and 98th percentile. Piloting revealed that some respondents struggle to think about risks, leading to unreasonably high (e.g. 0.5) or low (e.g. 0) risk estimates. Hence, we may winsorize to limit the presence of outliers.

6.1 VSL Estimate 1: Control, treatment 1, and treatment 2 data

This estimate is likely to have the best statistical power since it uses all available data. We will use all observations from the control and two treatment arms. The vector of treatment variables will include an indicator for assignment to treatment 1, an indicator for assignment to treatment 2, and the number of boda trips taken last week interacted with each of these variables.

6.2 VSL Estimate 2: Treatment 1 and treatment 2 data only

The second estimate of VSL will only use data from treatment 1 and treatment 2. We view this as the most credible estimate, particularly if priors are diffuse, since all material presented to the respondent is identical except for the estimated effectiveness of helmets. In particular, this estimate is robust to any endorsement effects that could come from presenting respondents with a study which generally shows that helmets are effective. This estimator may also reduce noise in beliefs since respondents are presented with empirical estimates of all risk variables. The limitation of

this approach is that the sample size is reduced.

6.3 Robustness to weak instruments

To account for the possibility of weak instruments, we plan to report first-stage F-statistics. In addition, we will report confidence sets robust to weak instruments. We will generally report confidence sets using the conditional likelihood ratio test from Moreira (2003). This test has optimal performance under homoskedasticity with one endogenous regressor. However, we may also report confidence sets constructed using alternative estimators, such as the heteroskedastic-robust AR from Chernozhukov and Hansen (2008). We currently do not plan to report AR test statistics as the primary specification since they lose power in over-identified models.

6.4 Robustness to endogenous future ridership

One concern with this study is that respondents could reduce their future ridership of motorcycle taxis in response to information about the risks associated with the mode of transportation or increase future ridership if they obtain a helmet. Based on discussions with the field team, we view this as unlikely. Furthermore, respondents are presented with the risk reduction offered by a helmet based on their past ridership, so this figure is more salient when providing valuations. However, to account for this possibility, we collect data about how many boda trips the respondent plans to take in the following week at the end of the survey. Hence, we can re-calculate the risk reduction, based on the respondent's own beliefs, using this ridership volume and re-estimate VSL for robustness.

6.5 Robustness to differing helmet lifespan

Respondents are told that helmets remain effective for 5 years, a number given to us by the helmet manufacturer. Furthermore, the 5 year lifespan is used to calculate the mortality risk reduction offered by a helmet. However, it is possible that some respondents may believe helmets have longer or shorter lifespans, changing the perceived likelihood that it will save their lives. We will thus collect data about how long respondents plan to use the helmet for. We will re-calculate risk reductions using the respondent's own beliefs about boda risks, helmet effectiveness, and their perceived lifespan of a helmet and re-estimate VSL for robustness.

6.6 The value of a statistical life year

In addition to VSL, we may estimate the value of a statistical life year (VSLY) using the same specifications, but considering the expected increase in life years associated with a helmet rather than the reduction in the probability that one will die as the covariate of interest.

References

Bachani, A. M., Y. W. Hung, S. Mogere, D. Akunga, J. Nyamari, and A. A. Hyder (2017, 3). Helmet wearing in Kenya: prevalence, knowledge, attitude, practice and implications. *Public Health 144*, S23–S31.

Becker, G. M., M. H. Degroot, and J. Marschak (1964). Measuring utility by a single-response sequential method. *Behavioral Science 9*(3), 226–232.

Belloni, A., V. Chernozhukov, and C. Hansen (2014, 4). Inference on Treatment Effects after Selection among High-Dimensional Controls. *The Review of Economic Studies 81*(2), 608–650.

Chernozhukov, V. and C. Hansen (2008, 7). The reduced form: A simple approach to inference with weak instruments. *Economics Letters 100*(1), 68–71.

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

Liu, B. C., R. Ivers, R. Norton, S. Boufous, S. Blows, and S. K. Lo (2008). Helmets for preventing injury in motorcycle riders. *The Cochrane database of systematic reviews* (1).

Moreira, M. J. (2003, 7). A Conditional Likelihood Ratio Test for Structural Models. *Econometrica 71*(4), 1027–1048.

Ouellet, J. V. and V. Kasantikul (2006, 3). Motorcycle Helmet Effect on a Per-Crash Basis in Thailand and the United States. *https://doi.org/10.1080/15389580500338652 7*(1), 49–54.

Romano, J. P. and M. Wolf (2005, 7). Stepwise Multiple Testing as Formalized Data Snooping. *Econometrica 73*(4), 1237–1282.

7 Appendix A: Survey instrument in English

VSL Survey

Field	Question	Answer	
Introduction			
intronote	Surveyor: Welcome to the boda safety survey. Please complete this section before speaking to the respondent.		
enumerator *(required)*	Select your name from the list.	1	Alex
		2	Francisca
		3	Julius
		4	Rachael
		5	Erick
		6	Susy
		7	Samson
		8	Christine
team *(required)*	Select your team.	1	Team 1 - CBD Region
		2	Team 2 - Eastlands Region
		3	Team 3 - Thika Road Region
		4	Team 4 - Ngong Region
constituency *(required)*	Select the constituency where you're working.	1	Westlands
		2	Kamukunji
		3	Starehe
		4	Mathare
		5	Embakasi South
		6	Embakasi North
		7	Embakasi Central
		8	Embakasi East
		9	Embakasi West
		10	Makadara
		11	Roysambu
		12	Kasarani
		13	Ruaraka
		14	Dagoretti North
		15	Dagoretti South
		16	Langata
		17	Kibra
terminal *(required)*	Select the boda terminal where the survey is taking place.	1	Yaya Centre
		2	Naivas Westlands
		3	Muthithi road
		4	Mpaka road
		5	MP Shah
		6	Loresho
		7	Kangemi
		8	Westgate
		9	Sarit Centre
		10	Thiongo
		11	Mountain View
		12	Rapta Road
		13	Riverside
		14	Parkland Mosque
		15	City park market
		16	Diamond Plaza
		17	Stima plaza
		18	Gikomba
		19	Majengo
		20	Kamukunji market
		21	Eastleigh garage
		22	Eastleighsection 3
		23	Eastleigh section 4
		24	St,Teresa
		25	Mlango Kubwa
		26	Muthurwa
		27	City Stadium

Field	Question	Answer	
		28	Kencom
		29	Hilton
		30	Koja
		31	GPO
		32	Nyamakima
		33	Railways
		34	Bus station
		35	Ngara/Fig tree
		36	OTC/Mosque
		37	Green Park
		38	Nyayo Stadium foot bridge
		39	KICC Fountain
		40	KNH
		41	Westlands Stage
		42	Kawangware
		43	Gikomba
		44	Chemist(Imara)
		45	River road
		46	Kirinyaga road
		47	Kariokor
		48	Luthuli Avenue
		49	Ambassador
		50	Accra Road
		51	Munyu Road
		52	Mathare Area 1
		53	Mathare Area2
		54	Mathare Area 3
		55	Mathare Area 4
		56	Mradi
		57	Number 10
		58	Naivas
		59	Stage 29/30
		60	Stage no. 30
		61	Huruma kwa chief
		62	Gateway
		63	JohnsangaKiamako
		64	Drive-inn stage
		65	Pipeline
		66	Tajmal
		67	Cooperative
		68	Transame
		69	Kobil
		70	Cabanas
		71	GM
		72	Entreprise Road
		73	Mukuru kwa njenga
		74	Imara Daima
		75	Makadara
		76	Makongeni
		77	Harambee
		78	Viwandani
		79	Maringo
		80	Hamza
		81	Mogas
		82	Uhuru stage
		83	Bahati stage
		84	Makadara Railway station
		85	Rikana
		86	Uchumi Embakasi North
		87	Riverside

Field	Question	Answer	
		88	River view
		89	Kwa chief
		90	Kariobangi
		91	Dandora 1 stage
		92	Dandora 2
		93	Dandora 3
		94	Dandora 4
		95	Kwa ndege
		96	Makuti
		97	Gate B
		98	Cousin
		99	Tassia Hill
		100	Quick Matt
		101	Kware
		102	Stage mpya
		103	Tassia Gate
		104	Donholm
		105	Tassia Complex
		106	Tassia stage
		107	Mowlem
		108	Kariobangi south kwa chief
		109	Stage 36
		110	Stage civo
		111	Outering stage
		112	Mutindwa
		113	Charina
		114	Umoja 1
		115	Umoja 2
		116	Lumumba Drive
		117	Mirema Stage
		118	Base Stage
		119	Canopy stage
		120	44 Stage
		121	Maziwa stage
		122	Kamiti stage
		123	Quickmart Kahawa stage
		124	Rounder Kahawawest stage
		125	Kahawa Wendani Stage
		126	Kahawa Sukari Stage
		127	K.U stage
		128	Isipe stage
		129	Equity stage
		130	Clay city stage
		131	Car wash stage
		132	Sunton Stage
		133	Maji Mazuri stage
		134	Mwiki stage
		135	Allsopes stage
		136	Naivas Stage
		137	Babadogo stage
		138	Lucky summer stage
		139	Riverside stage
		140	Maruirui stage
		141	Monetary Studies stage
		142	KCA University stage
		143	Pangani stage
		144	Muthaiga stage
		145	Lenana school drive
		146	Junction mall & Kingara rd junction

Field	Question	Answer	
		147	James Gichuru rd opp.Naivas
		148	Kawangware market
		149	Lavington Boarder
		150	Total-Corner
		151	Naivasha Road
		152	Shell petrol station
		153	Kuwinda RD
		154	Stedmark Garden
		155	The Well
		156	Queens Mall
		157	Langata Health centre
		158	Langata Market
		159	Southlands Estate
		160	Langata Cementry
		161	Langata Hospital
		162	Dawahab Hospital
		163	Langata Link road
		164	Wilson Airport
		165	Highrise flyover
		166	Kenyatta Market
		167	Ngong town opp. ABSA
		168	Ngong town bus stop
		169	Ngong Heights
		170	Zambia Stage
		171	Vet stage
		172	Embulbul opp. Rubis
		173	Bumps
		174	Kerarapon Drive opp. Maasai sch.
		175	Kerarapon Rd. opp shell
		176	Karen Shopping centre
		177	Kenyatta Market
		178	City Mortuary
		179	Prestige Junction
		180	U-turn
		181	D.C
		182	Makina Stage
		183	Karanja Stage
		184	Olympic stage
		185	Ayani-Equity
		186	42 stage
		187	Toy market
		188	Citam Woodley
date *(required)*	Please record the date and time.		

Field	Question	Answer	
resoindentnote	Surveyor: Now begin the interview with the respondent.		
welcome	My name is [enumerator_name] and I am working with the Principal Investigator (PI), Grady Killeen, a researcher from the University of California, Berkeley in the United States. We are conducting a study to better understand boda safety in Nairobi. This survey should take 15-45 minutes and participation is voluntary.		
consent *(required)*	Surveyor: Hand the printed consent form to the respondent (which they may keep for their records if they choose) and explain the form to them. Did the respondent sign the form and agree to participate in the survey?	1	Yes
		0	No
name *(required)*	What is your name? *Question relevant when: ${consent} = 1*		
age *(required)*	How old are you? *Enter -99 for refused.* *Question relevant when: ${consent} = 1* *Response constrained to: (.>17 and .<95) or .=-99*		
share_phone *(required)*	Are you willing to share your phone number so that we can call you if we have any questions about your responses? *Question relevant when: ${consent} = 1*	1	Yes
		0	No

Field	Question	Answer	
		.d	Don't know
		.r	Refuse to answer
phone *(required)*	What is your phone number? *Please enter the 10 digit phone number, beginning with 0.* *Question relevant when: ${share_phone} = 1* *Response constrained to: regex(., '^0[0-9]{9}$')*		
Additional respondent information *Group relevant when: ${consent} = 1*			
demographicsnote	Thank you. We would now like to ask you a few questions about yourself. These questions help us understand what types of people use bodas the most.		
gender *(required)*	Gender of the respondent.	0	Male
		1	Female
edusystem *(required)*	What school system did you participate in? *If more than one, select the most recent.*	.n	No schooling
		1	8-4-4.
		2	7-2-2-3
		3	2-6-6-3
		4	Other
		.d	Do not know
education *(required)*	What is your highest level of education? *Question relevant when: ${edusystem} = 1 or ${edusystem} = 2 or ${edusystem} = 3*	100	No schooling
		101	Std 1
		102	Std 2
		103	Std 3
		104	Std 4
		105	Std 5
		106	Std 6
		107	Std 7
		108	Std 8
		109	Form 1
		110	Form 2
		111	Form 3
		112	Form 4
		115	Some polytechnic
		116	Completed polytechnic
		117	Some college
		118	Completed college
		119	Some university
		120	Completed university
		121	Higher than college/ university
		122	Special education (mentally handicap)
		130	ECD/nursery/pre-unit
		200	No schooling
		201	Std 1
		202	Std 2
		203	Std 3
		204	Std 4
		205	Std 5
		206	Std 6
		207	Std 7
		209	Form 1
		210	Form 2
		211	Form 3
		212	Form 4
		213	Form 5
		214	Form 6
		215	Some polytechnic
		216	Completed polytechnic
		217	Some college
		218	Completed college

Field	Question	Answer
		219 Some university
		220 Completed university
		221 Higher than college/ university
		222 Special education (mentally handicap)
		230 ECD/nursery/pre-unit
		300 No schooling
		301 Pre-primary 1
		302 Pre-primary 2
		303 Grade 1
		304 Grade 2
		305 Grade 3
		306 Grade 4
		307 Grade 5
		308 Grade 6
		309 Junior Secondary 1
		310 Junior Secondary 2
		311 Junior Secondary 3
		312 Senior Secondary 1
		313 Senior Secondary 2
		314 Senior Secondary 3
		315 Some polytechnic
		316 Completed polytechnic
		317 Some college
		318 Completed college
		319 Some university
		320 Completed university
		321 Higher than college/ university
		322 Special education (mentally handicap)
		323 baby class/nursery/pre-unit
		.d Do not know
employed *(required)*	Are you currently employed? *If the respondent runs their own business, enter yes.*	1 Yes
		0 No
		.r Refuse to answer
wage *(required)*	Roughly how much do you earn in one hour of work? *Report the wage in Kenyan schillings. If the respondent has difficulty thinking about their hourly wage, enter -77. Enter -99 for refused to answer.* *Question relevant when: ${employed} = 1*	
monthly_wage *(required)*	How much did you earn last month? *Report earnings in Kenyan schillings. -77 for don't know* *Question relevant when: ${wage} = -77*	
hours_week *(required)*	How many hours do you spend working in a typical week? *-77 for don't know, -99 for refused* *Question relevant when: ${monthly_wage} > 0*	
future_wages *(required)*	Five years from now, how much do you expect to earn? *Enter the wages relative to the current wage. For instance, if they expect to earn twice as much in 5 years enter 2. Enter -77 for don't know and enter -99 for refused.* *Question relevant when: ${wage} > 0 or ${monthly_wage} > 0* *Response constrained to: (.>=0 and . < 101) or . = -77 or . = -99*	
children *(required)*	Do you have children?	1 Yes
		0 No
		.d Don't know
		.r Refuse to answer
health *(required)*	How would you describe your health these days? Would you say it is (read out):	1 Poor
		2 Fair
		3 Good
		4 Very good
		.r Refuse to answer

Field	Question	Answer	
life_expectancy *(required)*	Depending on environmental conditions and their individual health status, different people are expected to live for shorter or longer periods. Do you think you personally will still be alive at age ...? ENUMERATOR: Start asking at the nearest age above the age of the respondent. Stop asking as soon as the respondent mentions he/she will not be alive.	40	40
		45	45
		50	50
		55	55
		60	60
		65	65
		70	70
		75	75
		80	80
		85	85 or older
		.r	Refuse to answer
		.d	Don't know
recallnote	Thank you. We are now going to play a short game. I am now going to show you the tablet with several numbers written on it. I will show you the tablet for 10 seconds. I will then wait 10 seconds, and ask you to repeat the numbers back to me. Do you have any questions?		
digit1note	[digit1]		
digit1_ans *(required)*	Please repeat the number to me. *Surveyor: Enter the number exactly as the respondent states it. Enter -77 for don't know.*		
digit2note	[digit2] *Question relevant when: ${digit1_correct} = 1*		
digit2_ans *(required)*	Please repeat the number to me. *Surveyor: Enter the number exactly as the respondent states it. Enter -77 for don't know.* *Question relevant when: ${digit1_correct} = 1*		
digit3note	[digit3] *Question relevant when: ${digit2_correct} = 1*		
digit3_ans *(required)*	Please repeat the number to me. *Surveyor: Enter the number exactly as the respondent states it. Enter -77 for don't know.* *Question relevant when: ${digit2_correct} = 1*		
digit4note	[digit4] *Question relevant when: ${digit3_correct} = 1*		
digit4_ans *(required)*	Please repeat the number to me. *Surveyor: Enter the number exactly as the respondent states it. Enter -77 for don't know.* *Question relevant when: ${digit3_correct} = 1*		
digit5note	[digit5] *Question relevant when: ${digit4_correct} = 1*		
digit5_ans *(required)*	Please repeat the number to me. *Surveyor: Enter the number exactly as the respondent states it. Enter -77 for don't know.* *Question relevant when: ${digit4_correct} = 1*		
Boda use			
Group relevant when: ${consent} = 1			
boda_usenote	Thank you. I am now going to ask you several questions about your boda use.		
boda_uses *(required)*	Which of the following types of trips do you take a boda for?	1	Commute to/from work/school/church
		2	Purchase food
		3	Purchase other items
		4	Visit family
		5	Visit friends
		7	Deliveries (e.g. deliver goods)
		10	Emergency transportation/go to hospital
		6	Other
		.r	Refuse to answer
boda_uses_other *(required)*	Specify other boda uses *Question relevant when: selected(${boda_uses} , 6)*		
boda_trips *(required)*	In a typical week, about how many boda trips do you take? *Enter -77 for don't know*		
boda_trip_lenth *(required)*	How many minutes does a typical boda trip last? *Enter -77 for don't know*		
boda_reason *(required)*	Why do you typically choose to take a boda instead of another form of transportation? *Surveyor: Select all that apply.*	1	Bodas are fast
		2	Bodas are low cost
		3	Bodas are safe
		4	Bodas are the only option
		5	Bodas are fun

Field	Question	Answer
		7 Boda bodas are convenient
		8 Bodas are comfortable/private
		6 Other
		.d Don't know
		.r Refused to answer
boda_reason_other *(required)*	Please specify other reasons *Question relevant when: selected(${boda_reason} , 6)*	
Boda safety *Group relevant when: ${consent} = 1 and ${info_treatment} < 4*		
boda_safetynote	Thank you. We would now like to ask you questions about boda safety. We would like to better understand the risks that passengers face when they are not wearing a helmet. For the following questions, assume that passengers are NOT wearing helmets. Please think carefully about the following questions before responding. Please answer with your best guess even if you are not confident in your response. Some of these questions may be uncomfortable to think about. We are asking so that we can better understand the safety risks of boda bodas to help improve transportation safety, not to make you feel uncomfortable. We would be grateful if you are willing to respond.	
mortality_risk_initial *(required)*	We would like to better understand how often boda passengers are involved in fatal accidents. It is important for researchers to know how often fatal accidents occur in order to improve safety. For every 100,000 boda trips in Kenya, how many fatal accidents do you think occur on average? You may give a value less than one, such as 0.1 per 100,000, if you think that response is accurate. When answering, it may be useful to use the following values as reference points to help you express your answers. Jamhuri Day occurs 27 out of every 10,000 days or 270 out of every 100,000 days. A leap day, the 29th of February which occurs once every 4 years, occurs on 7 out of every 10,000 days or 70 out of every 100,000 days. So if you think that a fatal boda boda accident happens about 1/10th as often as a leap day, then there would be 7 fatal accidents per 100,000 trips. *Enter -77 for don't know or -99 for refuse.* *Response constrained to: . >=0 or . = -77 or . = -99*	
confidence_initial *(required)*	How confident are you that this value is correct? *Question relevant when: ${mortality_risk_initial} >= 0*	1 Very unconfident 2 Somewhat uncofident 3 Somewhat confident 4 Confident .d Don't know .r Refuse to answer
risk_initial_per_person	We would now like to understand the risk that a typical boda passenger faces. Consider passengers living in Nairobi that ride bodas as often as you do.	
risk_initial_1y_10k *(required)*	Over the course of 1 year, per 10,000 boda passengers how many do you think suffer a fatal accident? *Enter -77 for don't know or -99 for refuse.* *Response constrained to: . >=0 or . = -77 or . = -99*	
risk_initial_5y_10k *(required)*	Over the course of 5 years, per 10,000 boda passengers how many do you think suffer a fatal accident? *Enter -77 for don't know or -99 for refuse.* *Response constrained to: . >=0 or . = -77 or . = -99*	
confidence_initial_5y *(required)*	How confident are you that these values are correct?	1 Very unconfident 2 Somewhat uncofident 3 Somewhat confident 4 Confident .d Don't know .r Refuse to answer
boda_personalrisknote	Boda passengers' risk of suffering a fatal accident varies based on ridership, the routes that they take, and many other factors. We would now like to understand your personal risk of being in a fatal boda accident.	
boda_risk_5y_range *(required)*	In your view, what is the chance that you will be involved in a fatal boda accident over the next five years? *Surveyor: First select the range of values, then select the exact value on the next question.*	1 Greater than 1 in 10 (> 10%) 2 Between 1 in 10 and 1 in 100 (1%-10%) 3 Between 1 in 100 and 1 in 1000 (0.1%-1%) 4 Between 1 in 1000 and 1 in 10,000 (0.01%-.1%) 5 Between 1 in 10,000 and 1 in 100,000

Field	Question	Answer	
		6	Between 1 in 100,000 and 1 in 1,000,000
		7	Between 1 in 1,000,000 and 1 in 10,000,000
		8	Less than 1 in 10,000,000
		.r	Refuse to answer
boda_risk_5y *(required)*	In your view, what is the chance that you will be involved in a fatal boda accident over the next five years? *Question relevant when: ${boda_risk_5y_range} >= 0*	0.5	1 in 2 (50%)
		0.4	2 in 5 (40%)
		0.3	3 in 10 (30%)
		0.2	1 in 5 (20%)
		0.15	15 in 100 (15%)
		0.1	1 in 10 (10%)
		0.09	9 in 100 (9%)
		0.08	8 in 100 (8%)
		0.07	7 in 100 (7%)
		0.06	6 in 100 (6%)
		0.05	5 in 100 (5%)
		0.04	4 in 100 (4%)
		0.03	3 in 100 (3%)
		0.02	2 in 100 (2%)
		0.01	1 in 100 (1%)
		0.009	9 in 1,000 (0.9%)
		0.008	8 in 1,000 (0.8%)
		0.007	7 in 1,000 (0.7%)
		0.006	6 in 1,000 (0.6%)
		0.005	5 in 1,000 (0.5%)
		0.004	4 in 1,000 (0.4%)
		0.003	3 in 1,000 (0.3%)
		0.002	2 in 1,000 (0.2%)
		0.001	1 in 1,000 (0.1%)
		0.0009	9 in 10,000
		0.0008	8 in 10,000
		0.0007	7 in 10,000
		0.0006	6 in 10,000
		0.0005	5 in 10,000
		0.0004	4 in 10,000
		0.0003	3 in 10,000
		0.0002	2 in 10,000
		0.0001	1 in 10,000
		9e-05	9 in 100,000
		8e-05	8 in 100,000
		7e-05	7 in 100,000
		6e-05	6 in 100,000
		5e-05	5 in 100,000
		4e-05	4 in 100,000
		3e-05	3 in 100,000
		2e-05	2 in 100,000
		1e-05	1 in 100,000
		9e-06	9 in 1,000,000
		8e-06	8 in 1,000,000
		7e-06	7 in 1,000,000
		6e-06	6 in 1,000,000
		5e-06	5 in 1,000,000
		4e-06	4 in 1,000,000
		3e-06	3 in 1,000,000
		2e-06	2 in 1,000,000
		1e-06	1 in 1,000,000
		9e-07	9 in 10,000,000
		8e-07	8 in 10,000,000
		7e-07	7 in 10,000,000

Field	Question	Answer	
		6e-07	6 in 10,000,000
		5e-07	5 in 10,000,000
		4e-07	4 in 10,000,000
		3e-07	3 in 10,000,000
		2e-07	2 in 10,000,000
		1e-07	1 in 10,000,000
		1e-08	1 in 100,000,000
		0	0
previous_accident *(required)*	Have you been involved in a boda boda accident before?	1	Yes
		0	No
		.d	Don't know
		.r	Refuse to answer
contact_in_accident *(required)*	Do you personally know anyone that has been involved in a boda boda accident?	1	Yes
		0	No
		.d	Don't know
		.r	Refuse to answer
risk_source *(required)*	Which of the following sources of information did you consider when thinking of the risk of riding a boda?	1	My own experiences
		2	Information from friends or family members
		3	Information from the government
		4	News stories
		5	Information from social media (such as Facebook)
		6	Other
		.d	Don't know
		.r	Refused to answer
risk_source_other *(required)*	Specify other *Question relevant when: selected(${risk_source} , 6)*		
empirical_risk_note	We would next like to present you with estimates about how often passengers that ride bodas as often as you suffer fatal accidents. The estimates were calculated by the study team. We estimate that over the next 5 years, at least [empirical_boda_risk_display] out of every 10,000 passengers that ride bodas as frequently as you will die due to a boda accident. Equivalently, the risk that a passenger will die over the next 5 years is about [empirical_boda_risk_percent] percent or 1 in [empirical_boda_risk_rate]. These values use data from all of Kenya and not the Nairobi area only and the data may not record all fatal accidents, so they may underestimate your personal risk. *Sources: Daily trips from an op-ed by Fred Matiang'i (Secretary of the Ministry of the Interior) in Nation.Africa, mortality data from the National Transport and Safety Authority, and author's calculations.* *Question relevant when: ${info_treatment} > 1 and ${display_empirical_boda_risk} = 1*		
boda_risk_final_range	You estimated that the risk a typical Kenyan will suffer a fatal accident over the next 5 years is 1 in [initial_5y_rate] or [initial_5y_percent] percent. Your initial estimate of your own risk of suffering a fatal accident in the next 5 years was 1 in [personal_5y_rate] or [personal_5y_percent] percent. Based on the values you have heard, what is your final estimate of the risk of suffering a fatal boda accident over the next 5 years? Please ask if you would like to hear any of these values again.	1	Greater than 1 in 10 (> 10%)
		2	Between 1 in 10 and 1 in 100 (1%-10%)
		3	Between 1 in 100 and 1 in 1000 (0.1%-1%)
		4	Between 1 in 1000 and 1 in 10,000 (0.01%-.1%)
		5	Between 1 in 10,000 and 1 in 100,000
		6	Between 1 in 100,000 and 1 in 1,000,000
		7	Between 1 in 1,000,000 and 1 in 10,000,000
		8	Less than 1 in 10,000,000
		.r	Refuse to answer
boda_risk_final *(required)*	Enter the respondent's final estimate of their risk of suffering a fatal boda accident in the next 5 years. *Question relevant when: ${boda_risk_final_range} >= 0*	0.5	1 in 2 (50%)
		0.4	2 in 5 (40%)
		0.3	3 in 10 (30%)
		0.2	1 in 5 (20%)
		0.15	15 in 100 (15%)
		0.1	1 in 10 (10%)
		0.09	9 in 100 (9%)

Field	Question	Answer	
		0.08	8 in 100 (8%)
		0.07	7 in 100 (7%)
		0.06	6 in 100 (6%)
		0.05	5 in 100 (5%)
		0.04	4 in 100 (4%)
		0.03	3 in 100 (3%)
		0.02	2 in 100 (2%)
		0.01	1 in 100 (1%)
		0.009	9 in 1,000 (0.9%)
		0.008	8 in 1,000 (0.8%)
		0.007	7 in 1,000 (0.7%)
		0.006	6 in 1,000 (0.6%)
		0.005	5 in 1,000 (0.5%)
		0.004	4 in 1,000 (0.4%)
		0.003	3 in 1,000 (0.3%)
		0.002	2 in 1,000 (0.2%)
		0.001	1 in 1,000 (0.1%)
		0.0009	9 in 10,000
		0.0008	8 in 10,000
		0.0007	7 in 10,000
		0.0006	6 in 10,000
		0.0005	5 in 10,000
		0.0004	4 in 10,000
		0.0003	3 in 10,000
		0.0002	2 in 10,000
		0.0001	1 in 10,000
		9e-05	9 in 100,000
		8e-05	8 in 100,000
		7e-05	7 in 100,000
		6e-05	6 in 100,000
		5e-05	5 in 100,000
		4e-05	4 in 100,000
		3e-05	3 in 100,000
		2e-05	2 in 100,000
		1e-05	1 in 100,000
		9e-06	9 in 1,000,000
		8e-06	8 in 1,000,000
		7e-06	7 in 1,000,000
		6e-06	6 in 1,000,000
		5e-06	5 in 1,000,000
		4e-06	4 in 1,000,000
		3e-06	3 in 1,000,000
		2e-06	2 in 1,000,000
		1e-06	1 in 1,000,000
		9e-07	9 in 10,000,000
		8e-07	8 in 10,000,000
		7e-07	7 in 10,000,000
		6e-07	6 in 10,000,000
		5e-07	5 in 10,000,000
		4e-07	4 in 10,000,000
		3e-07	3 in 10,000,000
		2e-07	2 in 10,000,000
		1e-07	1 in 10,000,000
		1e-08	1 in 100,000,000
		0	0
helmetnote	Thank you. We would like to understand your views about boda helmets.		
helmet_info_treatment_2note	Academic studies show that high-quality motorcycle helmets can substantially reduce one's likelihood of dying in a boda accident. A 2008 review of studies on helmets estimates that they reduce the likelihood of a fatal accident by 42%. That is, the authors estimate that for every 100 individuals that would die in a boda accident without wearing a helmet, 42 people would survive if everyone wore a helmet. The studies that the authors examine look at countries		

Field	Question	Answer
	such as the United States where boda speed and helmet quality may differ from Kenya, so this study may not accurately reflect the effectiveness of using a helmet in Nairobi. Other studies may also produce different helmet effectiveness estimates. *Source: Liu BC, Ivers R, Norton R, Boufous S, Blows S, Lo SK. Helmets for preventing injury in motorcycle riders. Cochrane Database of Systematic Reviews 2008, Issue 1. Art. No.: CD004333. DOI: 10.1002/14651858.CD004333.pub3.* *Question relevant when: ${info_treatment} = 2*	
helmet_info_treatment_3note	Academic studies show that high-quality motorcycle helmets can substantially reduce one's likelihood of dying in a boda accident. A 2006 study using data from Thailand estimates that helmets reduce the likelihood of a fatal accident by about 70%. That is, the authors estimate that for every 100 individuals that would die in a boda accident without wearing a helmet, 70 people would survive if everyone wore a helmet. Boda speed and helmet quality may differ in Kenya, so this may not accurately reflect the effectiveness of using a helmet in Nairobi. Other studies may also produce different helmet effectiveness estimates. *Source: James V. Ouellet & Vira Kasantikul (2006) Motorcycle Helmet Effect on a Per-Crash Basis in Thailand and the United States, Traffic Injury Prevention, 7:1, 49-54, DOI: 10.1080/15389580500338652* *Question relevant when: ${info_treatment} = 3*	
helmet_effectiveness *(required)*	The effectiveness of boda helmets varies from individual to individual based on the types of trips that they take and other factors such as the speed of the operator. We would like to understand your own views about how effective boda helmets are. We would like to understand by what percent you think a helmet reduces a Kenyan's likelihood of dying in a boda accident. That is, for every 100 passengers that would die in a boda accident without wearing a helmet, how many do you think would survive if all 100 passengers wore a helmet? *Surveyor: For instance if the respondent thinks that helmets reduce one's chance of dying by 1/4, then enter 25.* *Response constrained to: (. >= 0 and . <= 100) or . = -77 or .=-99*	
helmet_effectiveness_confidence *(required)*	How confident are you that this value is correct? *Question relevant when: ${helmet_effectiveness} >= 0*	1 Very unconfident 2 Somewhat uncofident 3 Somewhat confident 4 Confident .d Don't know .r Refuse to answer
helmet_effectivenessnote	Thank you. The typical lifespan of a helmet is 5 years. Over this timeframe, this indicates that a helmet would decrease your chance of dying in a boda accident from 1 in [boda_risk_final_rate] to 1 in [risk_with_helmet]. In other words, for every 10,000 passengers like you, [risk_reduction] fewer people would die on average over the next 5 years if everyone wore a helmet compared to if no one wore a helmet.	
Willingness to pay		
Group relevant when: ${consent} = 1		
bdm_note	We would now like to play a short game to understand how much you would be willing to pay for a helmet. The helmet was produced by Boda Plus, a subsidiary of Car and General. The helmet adheres to the Kenyan Bureau of Standards' safety requirements. This requires passing a rigorous series of safety tests ensuring that the helmet will not easily break or fall off and that it absorbs impacts. The Kenyan Bureau of Standards requirements are similar to international standards created by the United Nations. You are going to have the option to receive either a free motorcycle helmet or a cash payment as a gift for participating in the study. The exact gift you receive will be determined by a game. We are first going to ask you what the smallest amount of money is that you would prefer to receive compared to a free helmet. In other words, what amount of cash do you value just as much as the helmet? We are then going to conduct a lottery in which we draw a cash payment amount at random. If the amount we draw is larger than the number that you state, then we will give you a payment in that amount. If the amount drawn is below the value that you state, then you will receive a free helmet. You cannot change your mind after we draw a random number. For example, if you stated KES 150 and the random draw was KES 149 or less, then you would receive a free helmet. But if the amount drawn was KES 150 or higher, then you would receive the cash payment. This lottery is designed so that it is always in your best interest to tell us exactly how much you value the helmet. The number that you tell us has no effect on the payment amount that will be drawn at random. Do you have any questions? *Note: If the respondent asks why everyone cannot receive the full payment, explain that the study budget does not permit us to offer everyone the largest payment.*	
wtp *(required)*	How much cash would make you just as well off as receiving a free helmet? *Response constrained to: .>= 0*	
confirm_valuation	You stated that you value a cash payment of [wtp] just as much as a free helmet. So if a payment of [wtp_below] or below is drawn, you would receive a free helmet. But if a payment of [wtp] or more is drawn then you would receive a cash payment. Is that correct?	

Field	Question	Answer
practice	We will now conduct a practice draw before the real one to help you understand how it will work. This will work the same as the real game, but we will not give you the item after this draw. The goal is to help you understand the game, and we will give you the opportunity to change your bid after the practice round.	
practice_offer_helmet	The practice draw is KES 500. Since this is below your bid, you would receive a free helmet if this were the real draw. *Question relevant when: ${helmet_practice} = 1*	
practice_offer_cash	The practice draw is KES 500. Since this is above your bid, you would receive a payment of KES 500 if this were the real draw. *Question relevant when: ${helmet_practice} = 0*	
final_confirmation *(required)*	Would you like to revise your bid before the final draw? It will NOT be possible to change your answer after this point. -	1 Yes 0 No
wtp_revised *(required)*	Surveyor: Enter the respondent's revised bid. *Question relevant when: ${final_confirmation} = 1* *Response constrained to: .>=0*	
offer_helmet	Thank you. The value selected is KES 160. You indicated you would prefer a free helmet compared to a cash payment in this amount. I will now help you select a helmet. *Question relevant when: ${helmet} = 1*	
nooffer_note	Thank you. You will receive KES 160. I will now arrange for you to receive the payment. *Surveyor: Explain how the respondent will receive the payment and answer any questions that they have.* *Question relevant when: ${helmet} = 0*	
helmet_delivered *(required)*	Surveyor: Did the respondent receive a helmet? *Question relevant when: ${helmet} = 1*	1 Yes 0 No
helmet_no_deliver_reason *(required)*	Why not? *Question relevant when: ${helmet_delivered} = 0*	
mpesa_number *(required)*	We will transfer the funds to you via the MPesa system. Which number should we send the funds to? *Please enter the 10 digit phone number, beginning with 0.* *Question relevant when: ${helmet} = 0* *Response constrained to: regex(., '^0[0-9]{9}$')*	
mpesa_number_confirm *(required)*	Please re-type the mobile number *Question relevant when: ${helmet} = 0* *Response constrained to: ${mpesa_number}_confirm = ${mpesa_number}*	
mpesa_provider *(required)*	What is the service provider of this SIM card for payment? *Question relevant when: ${helmet} = 0*	1 Safaricom 2 Airtel 3 Telkom .d Don't know
mpesa_name *(required)*	Please type the name as registered on the mobile account to receive the payment at *Question relevant when: ${helmet} = 0*	
mpesa_complete *(required)*	Surveyor: Were you able to complete the payment? *Question relevant when: ${helmet} = 0*	1 Yes 0 No .d Don't know .r Refuse to answer
mpesa_nocomplete_reason *(required)*	Why not? *Question relevant when: ${mpesa_complete} = 0*	
mpesa_nocomplete_note	Surveyor: Explain to the respondent what went wrong and notify them that they will receive the payment within 24 hours. *Question relevant when: ${mpesa_complete} = 0*	
nohelmet_reason *(required)*	Why did you not purchase a boda boda helmet prior to this survey? *Question relevant when: ${helmet} = 1*	1 I couldn't find an affordable helmet 2 I don't know where to buy a helmet 3 I couldn't find a high quality helmet 4 I never thought of purchasing a helmet 6 I expect the rider to provide a helmet 7 I don't own a boda boda/helmets come with boda boda purchases 5 Other .d Don't know

Field	Question	Answer	
nohelmet_reason_other *(required)*	Specify other *Question relevant when: selected(${nohelmet_reason} , 5)*		

helmet_lifespan *(required)*	For how many years do you plan to use the helmet?
	-77 for don't know, -99 for refused
	   Question relevant when: ${helmet} = 1
	   Response constrained to: . >= 0 or . = -77 or .=-99		

helmet_lifespan_nopurchase *(required)*	If you had received a helmet, how many years do you think you would you have used it for?
	-77 for don't know, -99 for refused
	   Question relevant when: ${helmet} = 0
	   Response constrained to: . >= 0 or . = -77 or .=-99		

planned_trips *(required)*	How many boda trips do you plan to take in the next week? *Response constrained to: . >= 0 or . = -77 or .=-99*		
Policy tradeoffs *Group relevant when: ${consent} = 1*			
policy_tradeoffs_note	Government officials often face tradeoffs between public safety and the cost of programs. For instance, the government may be able to build safer roads, but they would have to collect money from citizens like you to pay for the projects. We would like to better understand your views about these tradeoffs to help inform policy makers. There is no correct answer.		
support_covid_lockdown *(required)*	In 2020 and 2021, the Kenyan government imposed lockdowns to limit the spread of the covid-19 virus. Many people opposed these lockdowns because caused individuals to lose income from their jobs, but others supported them to protect public health. Do you oppose or support covid lockdowns?	0	Oppose
		1	Support
		.r	Don't know/refused
support_ambulances *(required)*	Suppose the government could offer ambulances to deliver individuals to hospitals during emergencies. Would you be willing to pay 1,250 KES every year to fund the service?	1	Yes
		0	No
		.d	Don't know
		.r	Refuse to answer
support_airbags *(required)*	Many government require airbags in vehicles to improve safety in the event of an accident. However, air bags cost about KES 25,000 on average. Should the Kenyan government require consumers to buy vehicles with air bags?	1	Yes
		0	No
		.d	Don't know
		.r	Refuse to answer
Search *Group relevant when: ${consent} = 1*			
search_note	Finally, we would like to ask you about where you could purchase a helmet if you wished to buy one. We hope to use this information to work with helmet producers to make them more accessible to consumers.		
know_helmet_seller *(required)*	If you wanted to purchase a high quality boda helmet, do you know the location of a seller where you could acquire one?	1	Yes
		0	No
		.d	Don't know
		.r	Refuse to answer
seller_location *(required)*	What is the location or contact information of the seller? *Surveyor: Enter as much detail as possible so we could find the seller later.* *Question relevant when: ${know_helmet_seller} =1*		
seller_search_time *(required)*	If you searched for a motorcycle helmet seller, how long do you think it would take to find one? *Surveyor: Ask the respondent to provide their best guess if they initially respond with don't know. Select the value closest to the respondent's answer.* *Question relevant when: ${know_helmet_seller} =0*	1	Half an hour or less
		2	Half an hour to an hour
		3	1 to 2 hours
		4	2 to 3 hours
		5	More than 3 hours
		6	I don't think I could find a helmet seller
		.r	Refused to answer
additional_seller_search_time *(required)*	If you wished to visit a second helmet seller to compare prices, how long do you think it would take to find one? *Surveyor: Ask the respondent to provide their best guess if they initially respond with don't know. Select the value closest to the respondent's answer.* *Question relevant when: ${know_helmet_seller} = 1 or ${know_helmet_seller} = 0*	1	Half an hour or less
		2	Half an hour to an hour
		3	1 to 2 hours
		4	2 to 3 hours
		5	More than 3 hours
		6	I don't think I could find a second helmet seller
		.r	Refused to answer
delivery_wtp *(required)*	Suppose that you wished to buy a motorcycle helmet. How much would you be willing to pay, in addition to the cost of the helmet, to have it delivered to you? *Surveyor: We want to know how much the respondent would be willing to pay in delivery fees to have the helmet delivered to their doorstep. This does NOT include the cost of the helmet.*	0	Nothing/would prefer to buy in a store
		25	KES 25
		50	KES 50
		100	KES 100

Field	Question	Answer	
		200	KES 200
		400	KES 400
		500	More than KES 400
		.d	Don't know/refused
endnote	Thank you for your participation. The survey is now complete.		
gps *(required)*	Please record a GPS point. *GPS coordinates can only be collected when outside.*		
consent_picture *(required)*	Surveyor: Take a picture of the signature page of the consent form. *Question relevant when: ${consent} = 1*		
complete *(required)*	Surveyor: Were you able to complete the survey?	1	Yes
		0	No
		.d	Don't know
		.r	Refuse to answer
nocomplete_reason *(required)*	Why not? *Question relevant when: ${complete} = 0*		