Are Estimates of the Value of a Statistical Life Unbiased?

Experimental and observational estimates in a low-income setting*

Grady Killeen

University of California, Berkeley

July 28, 2023

Abstract

A recent literature has estimated the value of a statistical life (VSL), individuals willingness to pay for a reduction in mortality risk, using revealed preference approaches in developing countries. Revealed preference estimates typically assume that individuals' beliefs about risk are equal to empirical estimates, and they assume that mortality risk is exogenous conditional on covariates. This paper reports the results of a study designed to experimentally estimate VSL and compare experimental to observational estimates. The experimental approach instruments for elicited beliefs using a randomly assigned information treatment, eliminating the need to use empirical risk estimates. I find that urban Kenyans have a low willingness to pay for safety. My preferred point estimate of VSL is just 2022 USD PPP $218, and I can reject values above USD PPP $420 with 95% confidence. Observational VSL estimates fall outside of experimental confidence sets, suggesting empirical risk estimates may be a poor proxy for beliefs. (*JEL* O18, R49, J17)

*I thank Edward Miguel for excellent feedback, advising, and help implementing this project. This project benefited from suggestions by Jeremy Magruder, Marco Gonzalez-Navarro, Sam Wang, Elif Tasar, Carlos Paramo, Catherine Che, and attendants of the UC Berkeley Development Lunch. I also thank William Jack, Whitney Tate, Nyambaga Muyesu, Josephine Okello and the team at the *Georgetown University Initiative on Innovation, Development, and Evaluation* for support with the implementation of the study. I gratefully acknowledge funding from the Center for Effective Global Action. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 2146752. This study received IRB approval from Amref Health Africa and the University of California, Berkeley. AEA Trial Registry RCT ID: AEARCTR-0010101. Email: gkilleen@berkeley.edu

1 Introduction

The value of a statistical life (VSL) – willingness to pay for reductions in mortality risk – has been studied extensively by economists because of the importance of mortality risk in a wide range of individual choices and public policies. The parameter is widely used by policy makers that face the problem of balancing lives saved from policies with their financial costs; academics that need to value reductions in mortality risk when studying decision making or optimal policy in risky environments; and, in the context of economic development, aid workers and donors that aim to direct funds to programs that maximize welfare. Despite the large empirical literature about VSL, the number of high quality studies from low-income settings is limited, and economists have recently raised concerns about the credibility of estimates even in high-income environments.

This paper produces precise experimental estimates of the VSL of urban Kenyans that are robust to prominent concerns that have been documented in the VSL literature. Furthermore, this study was designed to facilitate comparisons between experimental and observational estimates of the parameter across the same sample in order to provide the first test of the performance of commonly used methods. I find that consumers have a remarkably low VSL in this context: my preferred estimate rejects values above USD PPP $420 with 95% confidence. I further demonstrate that two revealed preference approaches to estimating VSL that have been used in recent publications produce point estimates that fall outside of experimental confidence sets and differ significantly from each other. Analysis of data on elicited mortality beliefs suggests that this discrepancy is driven at least in part by the fact that consumers' beliefs about the mortality risk of actions align poorly with empirical estimates, which are typically used as a proxy for beliefs.

A large number of studies have estimated VSL in high-income settings (for a review see

Banzhaf, 2014). Recent methods include examining demand for safety equipment (Rohlfs et al., 2015), comparing lives saved by public policies such as speed limits to economic costs (Ashenfelter and Greenstone, 2015), and estimating wage premiums for high risk jobs (Lee and Taylor, 2019).

Although there is a rich literature about VSL from rich countries, knowledge of demand for safety is limited in poor contexts. Common tools such as hedonic wage models depend on assumptions about perfect markets that are not credible in developing markets (Breza et al., 2021). The most credible studies of VSL in low income settings have thus turned to revealed preference approaches that estimate demand models for choices where fatality risk is salient (Kremer et al., 2011; León and Miguel, 2017; Berry et al., 2020; Ito and Zhang, 2020). However, Ashenfelter (2006) documents three methodological concerns with these studies.[1] First, risks may be endogenous, for instance due to selection into risky behaviors (Ashenfelter and Greenstone, 2004). Second, econometricians generally use empirical risk estimates as a proxy for beliefs, but agents may not have perfect information about risks. Third, VSL is likely heterogeneous across samples, and many studies estimate the parameter on selected populations.

This paper introduces an experimental approach to VSL estimation that makes progress on all three limitations of existing estimates. I first present motorcycle taxi passengers with randomly varying information about the fatality risk of motorcycles and the safety benefits of helmets. I then measure the posterior belief that a helmet will save the life of each respondent and elicit willingness to pay for a helmet using a Becker et al. (1964), hereafter referred to as BDM, mechanism. This design allows one to estimate VSL using treatment assignment as an instrument for elicited beliefs

[1] Ashenfelter (2006) also notes that agency problems raise concerns in some VSL estimates, but these do not apply to revealed preference estimates.

in a two-stage least squares regression. Random assignment addresses concerns of endogeneity and the use of elicited beliefs avoids the need to use empirical estimates as a proxy. The design of the experiment does not directly address external validity concerns, but the use of motorcycle taxis is near universal in this context and essentially no taxi passengers use helmets. Hence, the VSL estimates may be interpreted as an average over a broad urban sample, and I further show that VSL heterogeneity is limited in this setting.

My estimates of VSL indicate that urban Kenyan adults have low demand for safety. My preferred estimate is just 2022 PPP USD $218, and a weak-instrument robust confidence set rejects values greater than $420 with 95% confidence.[2] Two characteristics of VSL provide insight into why the parameter is so low in this context. First, VSL is a function of one's utility from being alive less one's expected utility from not being alive. The population in this study believes strongly in an afterlife, so VSL is likely much lower than one's utility from life. Second, VSL is inversely proportional to the marginal utility of a dollar, so theory predicts it will increase non-linearly with income (Hall and Jones, 2007). Given income differences between the sample in this study and León and Miguel (2017), which estimates a VSL (adjusted to 2022 terms) over $700,000, the two VSL estimates can be reconciled by a coefficient of relative risk aversion below 3.[3] These estimates have policy relevant welfare implications: they are an order of magnitude lower than VSL estimates used in 5 recently published benefit-cost analyses of health programs, causing the benefit-cost ratio to fall below 1 in 4 of 5 instances.

The study was designed to further facilitate comparisons between experimental and observational VSL estimates in order to evaluate the performance of revealed preference methods in

[2] I use the World Bank's 2021 PPP conversion factor to convert from Kenyan shillings to USD in this paper, which was 43.8 Kenyan shillings per USD at the time it was accessed.

[3] This study was selected for comparison because it reports incomes and is one of the highest revealed preference VSL incomes from a low-income country.

practice. I test two revealed preference approaches similar to those used in recent publications. First, I use cross-sectional variation in empirical mortality risk as a proxy for beliefs to estimate a demand model. This approach is similar to León and Miguel (2017) and Ito and Zhang (2020). Second, I use empirical mortality risk as a proxy for beliefs and assume that agents only value helmets for their life saving potential as in Berry et al. (2020) and Kremer et al. (2011). I find that the two approaches typically produce estimates that fall outside of experimental confidence sets, in some cases by orders of magnitude, and that they differ from each other by a statistically significant margin.

Data on beliefs indicate that empirical risk estimates are a poor proxy, helping explain these results. Respondents overwhelmingly reported using their own experiences and experiences of friends and family to generate beliefs, not empirical data sources. Empirical determinants of one's risk of dying in a motorcycle accident, such as the frequency with which they use them and the length of an average trip, are not positively correlated with beliefs. However, indicators for whether a respondent or someone they know has suffered a prior accident, which is not likely a function of luck after controlling for rich covariates, are strongly predictive of beliefs.

To my knowledge, this is the first study to produce a precise experimental estimate of VSL and the first to test the performance of observational approaches against an experimental benchmark.[4] This builds on recent revealed preference estimates of VSL from low-income countries, weakening identifying assumptions and estimating a VSL over a broad urban sample. This helps address a need identified in Greenstone and Jack (2015) for measures of willingness to pay for health among low-income populations. This study also contributes to the broader VSL literature

[4]Shrestha (2020) attempted to estimate VSL using experimental variation, but a weak first stage results in confidence intervals that are not informative.

by testing whether observational VSL estimates match an experimental benchmark. This provides among the first evidence about the degree to which the concerns identified in Ashenfelter (2006) bind in practice.

Aside from the estimates of VSL, this study contributes to a literature summarized in Delavande (2014) demonstrating that agents beliefs about stochastic events affect their decisions in low-income settings (e.g. Delavande, 2008; Attanasio and Kaufmann, 2017; McKenzie et al., 2013; Shrestha, 2020). I introduce a survey instrument that can elicit informative measurements of beliefs about low probability events. Furthermore, I show how the malleability of biased beliefs can be leveraged to identify parameters of a utility function.

Finally, the study contributes to a small literature about road safety. Traffic accidents are a leading cause of death in East Africa. Several studies have examined buses (e.g. Habyarimana and Jack, 2011, 2015). However, I am not aware of work studying motorcycle safety. I show that there is high demand for helmets despite low adoption, suggesting that improving accessibility of helmets could reduce traffic injuries and deaths.[5]

2 Study design

2.1 Sample: Motorcycle taxi passengers in Nairobi, Kenya

The sample of this study consists of motorcycle taxi passengers that did not own or regularly use a helmet prior to the experiment. The use of motorcycle taxis is common and rapidly growing in Kenya. There are an estimated 2.4 million drivers providing taxi services, combining for about 22

[5] High helmet demand exists despite low VSL primarily because agents have a high utility from characteristics of the helmet other than mortality risk reduction. Respondent anecdotally stated that they valued the protection from non-fatal injuries, which they viewed as costly due to lost work and medical expenses.

million trips per day.[6] Data from the National Transport and Safety Authority (NTSA) reports that 1,722 drivers and passengers died in 2021, compared to 715 deaths in 2017.

Motorcycle taxis represent an effective setting to study risk preferences because passengers face a significant mortality risk and rarely use helmets, which are effective at preventing death. Traffic deaths are the leading cause of death among boys 15-19 in Kenya, and a top-five cause of death for all Kenyans from 5-70.[7] Despite significant risks, helmet use among motorcycle taxi passengers has been measured at under 3% (Bachani et al., 2017). Given the efficacy of helmets, they are one of the most effective investments that individuals can make to reduce their mortality risk.

Low helmet use combined with the widespread use of motorcycle taxis allows for the estimation of VSL over a minimally selected sample. A concern is that consumers that select into using motorcycle taxis may have low valuations for safety given the risks of motorcycles. However, the use of motorcycle taxis is sufficiently widespread that it is unlikely that there is substantial selection. I am not aware of representative evidence documenting the prevalence of motorcycle taxi use, but the study team estimated that at least 85% of urban Kenyan adults use motorcycle taxis. Back of the envelope calculations also suggest that ridership needs to be near-universal among urban adults to rationalize total ridership volumes,[8] and alternative public transit options are also dangerous (Habyarimana and Jack, 2011).

[6] Fred Matiang'i, "The urgency of bodaboda reforms", *Nation.Africa*, 2022.

[7] "New initiative to tackle road crash deaths in Kenya," *World Health Organization*

[8] There were an estimated 22 million trips per day in Kenya in 2022, and a population of 53 million. 40% of Kenyans are farmers and at least 10% are children aged 5 or younger. Respondents in our sample report taking about 7.5 trips per week on average. This would suggest that over 90% of urban Kenyans adults ride motorcycle taxis.

2.2 Recruitment

This study recruited consumers from a motorcycle taxi stands in Nairobi during two waves of data collection. Surveyors censused a universe of 188 taxi stands and conducted surveys at 84, reaching 13 of Nairobi's 17 geographic constituencies. The stands were selected for broad geographic and demographic coverage. Dangerous locations were excluded for the safety of the field team. Survey locations are plotted over a map of Nairobi in Figure 1.

The study leveraged arrival times of consumers to sample from the population of the passengers at each location. Surveyors attempted to recruit the first individual to arrive at a taxi stand after completing a survey, and the high value of participation gifts relative to survey time yielded a high response rate. Over 90% of individuals approached agreed to take part in the survey. The majority of those that did not participate lacked the time to complete the survey, while a very small minority (under 1%) were excluded because they regularly used a helmet.

Demographic information presented in Table 1 shows that the study reached a broad demographic sample. This suggests that the VSL estimates are informative when considering other East African samples. Average education and income align closely with representative samples of the population. Over 70% of respondents reported completing secondary school, whereas the World Bank reported a lower secondary school completion rate of 79.2% for Kenya in 2016. The mean annual income of control respondents is USD PPP $6,730 with a median of $4,762. The World Bank reported a GDP/capita of $5,211 for Kenya as a whole in 2021, and the Kenyan National Bureau of Statistics reported gross per capita production of $7,907 for Nairobi county in 2017.[9] The sample is not, however, perfectly representative. Significantly more males than females were surveyed, explained by the fact that that men are more likely to commute.

[9] Source: 2019 Gross County Product Report and 2017 World Bank PPP conversion rate.

2.3 Information treatment: Motorcycle fatality risks and helmet effectiveness

This study implemented a randomized information treatment to produce the variation in beliefs needed to estimate VSL and examine whether biased beliefs explain low helmet use. There are four experimental arms: a pure control group, a control group, and two treatment arms. The two treatment arms were presented with information about the mortality risk of motorcycle taxis in Kenya and the effectiveness of helmets at preventing death. The two control arms received no information.

The pure control and control arms vary in the survey questions that they were asked. The pure control was not asked any questions about motorcycle safety. In contrast, the control group was asked detailed questions about the risks of using motorcycle taxis and surveyors elicited beliefs about the effectiveness of helmets. The pure control group was included because I hypothesized that asking respondents about the risks of motorcycles and the benefits of helmets may increase helmet demand, changing the public policy implications of results. Pure control observations cannot be used for VSL estimation because beliefs are not elicited. In practice, there were no differences in helmet demand between the pure control group and control group during the first wave, so the pure control arm was not included in the second wave of the study.

The two treatment arms were presented different studies estimating the effectiveness of motorcycle helmets. All respondents in these treatment arms first received an estimate of their baseline empirical mortality risk over the 5 year lifespan of a helmet, estimated based on their ridership volume using data provided by the NTSA.[10] Respondents were then presented with one of two studies about the effectiveness of helmets at preventing death. Those in a “low treatment” group

[10] I calculated per trip mortality risk estimates for the average Kenyan from NTSA data, then estimated 5 year risks based on the respondent’s expected ridership volume over that period.

were presented with the results of Liu et al. (2008) which reports the results of a meta-analysis of studies of helmet effectiveness. The authors estimate that helmets reduce one's likelihood of dying by 42%. Respondents assigned to a "high treatment" arm were presented with the results of Ouellet and Kasantikul (2006), who estimate that high quality helmets reduce fatality risk by about 70% in Thailand.

Both studies are credible and no respondents were given misleading information about helmet efficacy. There is a strong consensus that motorcycle helmets are effective, but there is considerable uncertainty about exactly how well they work. In fact, the 70% estimate reported in Ouellet and Kasantikul (2006) is approximately the upper bound of the 95% confidence interval from Liu et al. (2008). Furthermore, there is a trade-off between quality and contextual similarity between the studies. Liu et al. (2008) is a high-quality meta analysis, but the authors mainly consider studies from rich countries where roads quality and traffic speeds are higher. Ouellet and Kasantikul (2006) reports results from a single study, but the authors examine a low income setting.

2.4 Helmet valuations

The study measured demand for a high quality helmet using a Becker et al. (1964) willingness to accept mechanism. Demand was elicited after delivering the randomized information treatments to measure the effect of the information on willingness to pay. Respondents were asked to state the smallest cash payment that they would prefer to a free helmet. Surveyors then revealed a randomly selected payment amount, drawn from a uniform distribution between Ksh 0 and 600 and rounded to the nearest 10 shillings. If the payment amount was greater than or equal to the respondent's bid, then they received the cash payment. Otherwise they were given the free helmet. The study used a willingness to accept rather than a willingness to pay mechanism to ensure that liquidity constraints did not bind, which would result in underestimating VSL.

The maximum draw was set to Ksh 600 based on assessments by experts at an NGO and the helmet manufacturer that most valuations would fall below this threshold. The helmet model used in the study sold at a wholesale price of 580 Kenyan shillings (Ksh), about 14% of weekly wages for the median respondent. A limitation of the willingness to accept mechanism is that disclosing one's true valuation is not the unique weakly dominant strategy if the valuation exceeds the maximum draw. To hedge against the risk of setting the max payout too low, and to avoid anchoring effects, enumerators did not disclose the upper bound or value of the helmet when introducing the game.[11]

In practice, helmet valuations exceeded the wholesale price of the helmets in about $3/4$ of cases. Figure 2 demonstrates that there is no unusual behavior in the distribution of bids near the maximum BDM draw, consistent with accurate valuation data. Panel A plots a histogram dropping bids over Ksh 2,000, which does not result in any excess or missing mass near 600, and Panel B plots the full distribution of bids.

2.5 Randomization

Respondents were assigned to the four information arms using a pseudo random number drawn using the survey software. Randomization was not stratified since the sample was not defined in advance. During the first wave, respondents were assigned to the pure control group with a probability of 0.1 and each of the other three groups with a probability of 0.3. During the second wave, the pure control group was eliminated and respondents were assigned to the three remaining arms with equal probability. An independent random draw determined the cash payment amount

[11] Surveyors were instructed to reveal this information if asked about it, but there was only a single report of this happening.

offered to each respondent.

3 Model

This section presents a simple model of demand for motorcycle helmets. This demonstrates how the experiment identifies VSL and shows that it may be estimated using a two-stage least squares regression. Furthermore, the model explains how most existing revealed preference VSL estimates are calculated and demonstrates that these estimates are unbiased only in the knife-edge case in which all individuals in the sample have mortality beliefs perfectly aligned with the econometrician's estimate of the empirical risk of a decision.

Consider a set of individuals indexed by $i \in \{1, \ldots, I\}$. I assume that individuals are expected utility maximizers and that their beliefs about motorcycle mortality risk and helmet effectiveness evolve according to a Bayesian learning process. The assumption that agents maximize expected utility is necessary to identify VSL. The assumption that agents Bayesian update is not. Although this assumption is reasonable, this process is chosen primarily for convenience and to illustrate how micro-founded learning models can generate bias in cross-sectional estimates that use empirical mortality risk as a proxy for beliefs. To identify VSL, I require the weaker condition that the information presented in the study leads to a change in beliefs, which I test empirically.

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

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

where $\mathcal{I}_0$ denotes the individual's baseline information set. I temporarily exclude the subscript i

for clarity, but I do not assume that beliefs are constant across individuals. Taking the expectation

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

Now suppose that the consumer receives a signal, in this case from the surveyor, 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's posterior mean will differ from their prior. 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. If consumers with equal priors are given two different signals about the efficacy of helmets, in this case from the two different studies used in the two treatment arms, then their posterior means will differ so long as their prior is non-degenerate.

Similarly, suppose that the agent has a prior distribution about the probability of getting into a fatal motorcycle accident without a helmet, per trip, given by

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

After receiving a signal about the empirical accident risk is $\theta_A \sim Beta(\alpha_{EA}, \beta_{EA})$ their posterior mean will be

$$A_1 = \frac{\alpha_{0A} + \alpha_{EA}}{\alpha_{0A} + \alpha_{EA} + \beta_{0A} + \beta_{EA}}$$

If agents learn about risk through their own experiences or their social network then their beliefs will likely vary from empirical estimates of risk. If agents learn from their own experiences, then due to selection on survivorship their posterior belief about the per trip probability of dying will converge to zero. This would generate an ambiguous relationship between ridership volume, which empirically increases mortality risk.

Learning from one's social network will also produce biased beliefs even if agents are equally likely to hear about fatal and non-fatal journeys because accidents are rare. Based on NTSA data, I estimate the chance that an individual with the median ridership volume (6 trips per week) will die in a motorcycle accident over a 5 year period is about 1 in 4,000. If individuals observe the transportation outcomes of 100 friends on average, then in expectation 0.025 people in the agent's social network would die over 5 years. Since empirically the agent can only observe a discrete number of people dying, then consumers would under or overestimate risk (an instance of hasty generalization as documented in Rabin, 2002). Differential probabilities of learning about fatal vs non-fatal trips would produce further bias.

Suppose the consumer completes n motorcycle taxi rides over the lifespan of a helmet. Their baseline expectation of mortality risk over this time frame without a helmet is given by

$$r_{in} = 1 - \int_{A=0}^{1} (1 - A)^{n_i} Pr(A|\mathcal{I}_0)$$

Under prior beliefs with a helmet, it is given by

$$r_{ih0} = 1 - \int_{A=0}^{1} \int_{D=0}^{1} (1 - (A \cdot D))^{n_i} Pr(A|\mathcal{I}_0) Pr(D|H; \mathcal{I}_0)$$

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

$$r_{ih1} = 1 - \int_{A=0}^{1} \int_{D=0}^{1} (1 - (A \cdot D))^{n_i} Pr(A|\mathcal{I}_0) Pr(D|H; \mathcal{I}_1)$$

Letting p_i denote the price of a helmet and r_{ih}, r_{in} denote the consumer's belief about their risk with and without a helmet respectively. Let $\beta = u(\text{alive}) - u(\text{not alive})$ denote the difference in one's utility from being alive versus not alive. It is important to note that one's expected utility from not being alive need not be zero, so this does not simplify to the present value of lifetime utility. In fact, the population in this study has strong views of an afterlife, so we would expect $u(\text{not alive}) > 0$. The consumer's expected utility from purchasing a helmet is thus

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

where $v = \frac{\beta}{\alpha}$ is the value of a statistical life and ϵ_{ih} represents components of utility unobserved to the econometrician.[12] ζ_h captures the average utility that agents receive from characteristics of a helmet other than its life-saving potential, such as protection against injuries or discomfort. Without a helmet, the consumer's expected utility is

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

[12] I assume that the marginal utility of a dollar, α, and the difference in utility from being alive versus not β are constant for simplicity. Of course, both parameters likely vary with factors such as wealth.

Let $y_i = 1$ if the consumer purchases a helmet, which corresponds to $U_{ih} \geq U_{in}$, and denote $\Delta r_i = r_{in} - r_{ih}$. Normalizing $\zeta_n = 0$ and assuming that $\epsilon_i = \epsilon_{ih} - \epsilon_{in} \sim \mathcal{N}(0, 1)$ produces the probit model

$$Pr(y_i = 1|\mathcal{I}) = \Phi\left(\zeta_h + \beta\Delta r_i - \alpha p_i\right)$$

To estimate VSL, researchers may use data on y_i, r_{in}, r_{ih} and p_i to estimate the demand parameters α and β. The value of a statistical life, that is their willingness to pay to reduce mortality risk, is then given by $VSL = \frac{\beta}{\alpha}$.

A typical assumption in the VSL literature is that an individual's belief about mortality risk is equivalent to an estimate of empirical risk calculated by the econometrician. In this context, this assumption would be that $\Delta r_i = \Delta r_i^*$ for all i where Δr_i^* is the empirical reduction in mortality risk associated with a helmet, conditional on the agent's ridership characteristics. However, this will return a biased estimate of VSL unless beliefs perfectly align with the estimated risk.[13]

This assumption has been made because of a lack of alternative methodology capable of producing meaningful VSL estimates.[14] Mortality beliefs are generally measured with noise, so simply including stated expectations in regressions would attenuate estimates.

This study avoids the need to use empirical risk estimates by experimentally generating an

[13]Even if empirical beliefs and calculated risks align on average, but not for all individuals, there will be attenuation bias.

[14]For instance, León and Miguel (2017) note that "we follow the existing literature and utilize a standard expected utility individual choice framework, using accident risk from historical data, in part due to the absence of a well-articulated and widely accepted alternative analytical approach that incorporates these behavioral concerns and generates meaningful valuation estimates."

instrument for beliefs. Let v_i denote individual i's willingness to pay for a helmet. Then

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

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

$$v_i = \frac{1}{\alpha}\zeta_h + VSL\Delta r_i + \frac{1}{\alpha}\epsilon_i$$

Hence, one may identify VSL using a simple two-stage least squares regression instrumenting for Δr_i with treatment assignment. This approach does not rely on distributional assumptions about unobserved determinants of utility.

Leveraging data on expected remaining lifespan, LS_i, then

$$v_i = \frac{1}{\alpha}\zeta_h + VSLY\Delta r_i \cdot LS_i + \frac{1}{\alpha}\epsilon_i$$

So the same approach may be used to construct revealed preference estimates of the value of a statistical life year (VSLY).

4 Data and empirical specification

4.1 Data

I use data from 1,571 surveys completed in two waves. The first wave was conducted between October and December 2022, and the second between February and March 2023. The first wave consisted of a total of 921 surveys, including pure control observations, and the second wave included 650 surveys.

The survey collected detailed demographic data and information about the motorcycle taxi use of all respondents, and baseline and posterior beliefs about motorcycle taxi risks from those that were not assigned to the pure control group. The methodology used to elicit beliefs was refined during a pilot exercise. Details of the final approach are presented in Appendix A and the full instrument is available on the AEA RCT Registry. The survey elicits priors about the per trip mortality risk of a motorcycle taxi, the average number of deaths per 10,000 motorcycle taxi passengers over 1 and 5 year periods, and the respondent's risk of dying in a motorcycle accident over 5 years. Posterior beliefs about the respondent's 5 year mortality risk without a helmet and the effectiveness of helmets at preventing death were collected. Multiple measures of baseline beliefs were collected to validate the variables against each other.

Table 1 presents demographic characteristics of the sample and demonstrates balance across treatment arms with respect to these variables. Table 2 is similar, but it presents summary statistics with respect to motorcycle taxi use, reasons for motorcycle taxi use, priors about their riskiness, and sources of information used to construct beliefs about risk by treatment arm. These outcomes are also generally balanced across experimental groups, although there is some imbalance between the pure control and other arms. Appendix Table A1 and Appendix Table A2 demonstrate that non-response rates are similar across treatment arms and that response rates are high.[15]

[15]Despite the sensitive nature of questions about one's mortality risk, only about 2% of respondents declined to answer.

4.2 Experimental VSL estimation

My primary estimate of the value of a statistical life is obtained via the two-stage least squares regression model

$$v_i = \zeta_h + VSL\Delta r_i + X_i'\gamma_0 + \gamma_1 r_{0,i} + \epsilon_i$$

$$\Delta r_i = Z_i'\pi + X_i'\pi_c + \pi_r r_{0,i} + \nu_i$$

where v_i is the respondent's willingness to pay for a helmet, Δr_i is the reduction in mortality risk that a respondent perceives over the 5 year lifespan of a helmet, X_i is a vector of controls, and $r_{0,i}$ is the respondent's baseline belief about their 5-year mortality risk without a helmet. Z_i denotes a vector of excluded instruments. I report results with two sets of instruments. First, I consider an "interacted" set of instruments $Z_i = (T_i', r_{0,i} \cdot T_i')'$ where T_i is a vector of treatment assignment indicators. The second set of instruments consists of treatment assignment only, that is $Z_i = T_i$.

My preferred instrument is the interacted version because it absorbs variation in baseline heterogeneity in beliefs, leading to substantial power gains if heterogeneity in VSL is small or orthogonal to $r_{0,i}$. Intuitively, if there is heterogeneity in the perceived risk of riding motorcycle taxis across the population, then Δr_i will vary both due to differences in risk and beliefs about the efficacy of helmets. If heterogeneity in perceived motorcycle accident risk is large then VSL will be less precisely estimated using treatment assignment only than the interacted instrument that accounts for priors. [16]

[16] I pre-specified two sets of instruments before the first wave of the experiment, but the interacted version varied from what we use in the paper. The original PAP specified the use of n_i, the number of motorcycle taxi trips taken in a typical week, in place of $r_{0,i}$. The logic was the same, but I made the assumption that n_i, which is important in determining empirical risk, would be an effective proxy for beliefs. Baseline beliefs were in fact orthogonal to n_i during wave 1, so it does not help account for heterogeneity. Hence, I filed a PAP amendment before wave 2 to instead

Controls were selected using single-post LASSO. The set of possible controls includes demographic variables, motorcycle trip types and reason for use, and the information sources used to construct beliefs about mortality risk.[17] Estimates also include surveyor fixed effects because surveyors were assigned to fixed parts of the city and some field officers were more effective at answering questions about the studies, essentially leading to a stronger information treatment.[18]

I follow the PAP and report results over two samples. First, I use data from the control and both treatment arms. Second, I restrict the sample to treated respondents. The restricted sample is presented to demonstrate that results are not driven by an endorsement effect associated with informing individuals that academic studies show helmets are effective.[19]

4.3 Observational VSL estimation procedures for comparison

I implement two observational approaches to estimating VSL similar to those used in recent publications. First, I estimate the OLS regression

$$v_i = \zeta_h + VSL\Delta r_i^* + X_i'\gamma_0 + \epsilon_i$$

where Δr_i^* is the empirical likelihood that a helmet will save the respondent's life, estimated from ridership and treating the point estimate from one of the two helmet studies as true. This approach is similar to León and Miguel (2017) and Ito and Zhang (2020) which estimate logit demand models using empirical risk estimates as a proxy for beliefs. I similarly estimate a demand using

use $r_{0,i}$. The treatment only model is unmodified from the original PAP.

[17] Due to an error, the initial PAP only listed the demographic variables. The PAP amendment filed before wave 2 specified the full set of potential controls.

[18] These were added in the PAP amendment. Results are similar if they are excluded.

[19] I report homoskedastic standard errors and weak instrument robust confidence sets in the primary tables and two-stage cluster bootstrapped standard errors and confidence intervals (Abadie et al., 2022) in the appendix. Appendix B provides details about inference, including justifications for these decisions.

empirical estimates of the mortality risk from a helmet, but the specification is simpler since I have willingness to pay data. Following León and Miguel (2017), I do not instrument for empirical risk.

Second, I report the estimates

$$VSL_i = \frac{v_i}{\Delta r_i^*}$$

This approach assumes that agents do not receive utility from any characteristics of helmets other than their life saving potential and that $\Delta r_i = \Delta r_i^*$ for all i. This specification is identical to Berry et al. (2020) and is similar to Kremer et al. (2011).

5 Results

5.1 Validating belief measurement

Elicited beliefs about the likelihood of suffering a fatal motorcycle accident fall in plausible ranges and are strongly correlated across measurements. This suggests that the data capture an informative signal about respondents' true beliefs.

The mean baseline five year mortality risk is 0.034, the median is .001, the 1st percentile is 10^{-7} and the 99th percentile is 0.5. I estimated that the median empirical risk was about 1 in 4,000, but the NTSA data used to calculate this may not have accounted for unreported accidents. Responses about the number of deaths per 10,000 passengers per 5 years and the assessments of one's own risk were highly correlated ($p < .01$, $R^2 > 0.1$), although the median consumer believed that they were only about 20% as likely to die as an average Kenyan. This would be consistent with overconfidence in one's own safety.

Beliefs about the 5 year and 1 year mortality rates of motorcycle passengers are strongly correlated. The R^2 between the measures exceeds 0.57, and for over 60% of respondents, the 5 year estimate is between 4 and 6 times their 1 year estimate. On average, the 5 year response was 5.24 times the 1 year response. Only 15 respondents stated that they believed more people would die over the next 1 year than the next 5 years, which is impossible. The respondents' belief about the 5 year risk of an average Kenyan were also positively related with their beliefs about the per trip risk of a fatal accident ($p < .01, R^2 = 0.2$).

The right tail of responses raises a concern that some respondents may have trouble understanding or communicating probabilities. It is striking that one would choose to ride motorcycle taxis if they felt that the risk of dying is 50% over 5 years. However, respondents that gave very high responses reported them consistently high across measurement approaches. Surveyors were also instructed to ask, after the survey was complete, why they chose to ride motorcycle taxis if they felt the risk was so high. Typical responses were that they had been told that motorcycles were extremely dangerous, which they interpreted as a value such as 1 in 2. When asked why they still chose to use them, respondents indicated that they either had no safer way of commuting for work or that they were content with the possibility of dying due to religious beliefs. Hence, these extreme responses seem to reflect very high assessed probabilities of suffering a fatal accident, rather than difficulties understanding probabilities.

5.2 Correlates with beliefs

Respondents reported forming beliefs primarily based on their own experiences and information from their social network. Consistent with these accounts, I find that variables indicating whether the respondent or someone in their social network suffered a prior accident are strongly correlated with beliefs. This provides further evidence that belief data is informative.

Figure 3 plots the information sources that respondents reported using to form their beliefs about the risks of motorcycle taxis. The most common information source reported is passengers' own experiences (79%), the second most common information source considered is the experience of friends or family members (48%), and the third most common source is social media (38%). Over 95% of respondents stated that they considered their own or others' experiences. In contrast, under a third of respondents reported that media or government sources, which are likely more objective, informed their views about safety. Beliefs should therefore be correlated with the accident history of oneself and their social network. Furthermore, the fact that consumers overwhelmingly form their beliefs based on personal experiences suggests that they are unlikely to coincide with empirical estimates.[20] This suggests that beliefs need not be correlated with empirical determinants of accident risk.

Table 3 confirms these predictions. In columns 1-2, I estimate the least squares fit of one's prior belief of dying in a motorcycle accident on ridership volume. The correlation is negative in both column 1, which excludes controls, and in column 2 which includes taxi terminal fixed effects and covariates selected using double-post LASSO (Belloni et al., 2014). This suggests that agents do not learn from unbiased sources since empirically those that ride motorcycles more often are more likely to die on them, especially controlling for the taxi stand where the survey was conducted at since the set of drivers and routes is then small. Similarly, columns 3-4 show that the length of one's average trip, which one would again expect to be correlated with unbiased estimates of risk, is not strongly correlated with beliefs.

Columns 5-8 examine the correlation between beliefs and an indicator for whether the respondent suffered a prior motorcycle accident and an indicator for whether a member of the respon-

[20] As noted in Section 3, even if the agent's social network is perfectly representative and survived trips are just as likely to be reported as deaths, accidents are sufficiently rare that agents forming beliefs based on their social network's experiences are likely to be biased.

dent's social network suffered a motorcycle accident. Consistent with stated information sources, those that suffered a prior accident perceive higher risks of using motorcycle taxis, although this relationship is not statistically significant once controls and terminal fixed effects are added. Individuals that know someone that has experienced an accident also perceive motorcycle taxis as significantly more dangerous, and this relationship survives the addition of controls.

5.3 First stage effects on beliefs

Table 4 demonstrates that randomized information exposure had a statistically significant effect on the agent's posterior belief that a motorcycle helmet will save their life. This implies that information treatment assignment may be used as a valid instrument to identify VSL.

I first examine the effect of receiving an estimate of one's empirical likelihood of dying in a motorcycle accident without a helmet on beliefs in columns 1 and 2. Respondents did not update their beliefs based on this information. There is not a significant difference between either treatment arm, who received identical information, and the control group. Surveyors reported that respondents understood the information being presented, but they did not trust the estimates because of a view that the NTSA is ineffective at reporting accidents. The fact that respondents did not report updated beliefs in response to this information, and that they felt confident telling surveyors that they did not trust the information, provides evidence that surveyor demand effects are not present in this experiment.

Respondents did update their beliefs about the effectiveness of helmets based on the studies presented in the low and high treatments. Table 4 reports treatment effects on beliefs about the effectiveness of helmets in columns 3 and 4. The mean belief about helmet effectiveness reported in the control group was about 79%. This value exceeds most academic estimates, and many control respondents stated effectiveness beliefs over 90%. The average in the high treatment group was about 75% by contrast, and that in the low treatment group was about 65%. The differences between the treatments arms are significant at the 1% level.

Similarly to the information on accident risks, during piloting and the start of data collection respondents' beliefs about the effectiveness of helmets did not significantly update. Surveyors reported that respondents were asking follow-up questions about the studies that they did not know how to answer, such as inquiring about causes of death observed when helmets were used. Difficulty responding to these questions caused respondents to doubt the credibility of the information, and they reported that they did not trust the quality of the studies. Hence, surveyors were trained with details about the two studies so that they could respond to these questions. After this additional training, the information treatments began to affect beliefs and average posteriors fell below the study estimates and the control average posterior. The fact that respondents did not update beliefs until surveyors could provide adequate details about the studies and that posteriors exceeded study estimates on average provides further evidence that surveyor demand effects are not binding in this study.

Despite a lack of updating about motorcycle accident risks, the change in beliefs about helmet effectiveness were sufficient to detect a change in the probability that a helmet would save the respondent's life. Columns 5 - 6 of Table 4 report the change in the perceived lives saved per 10,000 people associated with helmet use over 5 years. Beliefs about accident risk and helmet effectiveness among those assigned to the low treatment arm correspond to around 80 fewer lives saved on average compared to the control group. This effect is driven in part by outlying beliefs: the difference falls to about 46 fewer lives if data is winsorized at the 2nd and 98th percentiles. These differences are significantly different from the control and high treatment beliefs at the 1%

level. There is not a significant difference between the high treatment and control means, although point estimates change in the expected direction. Overall, the variation in beliefs induced by the treatment creates sufficient variation to estimate the value of a statistical life.

5.4 The value of a statistical life

Table 5 demonstrates that demand for safety is low in this sample. My preferred estimate of the value of a statistical life uses the interacted set of instruments and controls selected using LASSO. This value is reported in column 3 of panel a. The point estimate of the value of a statistical life is just USD PPP $219, and a weak instrument robust confidence set rejects values below $34 and above $420 with 95% confidence. Results are similar without controls, using the non-interacted instrument, and if willingness to pay data is winsorized. However, there is a drop in power in these specifications, meaning that they do not always exclude 0.

Panel b shows that results are similar if the sample is restricted to those that were treated. These results are robust to potential confounds associated with an endorsement effect driven by presenting respondents with rigorous data showing that helmets are effective, but there is a loss in power since $1/3$ of the sample is excluded. Despite the loss in power, one may reject VSL estimates above $1,006 with 95% confidence using the preferred specification and above $2,524 across all specifications.

The estimates in Table 5 exclude 33 observations that the field team reported were contaminated by motorcycle drivers pretending to be passengers and submitting false survey responses in order to receive a free helmet.[21] Appendix Table A3 reports first stage results excluding these observations

[21] This occurred on the second day of data collection because teams returned to an area that they had already visited and motorcycle drivers recognized them. I exclude all observations from this day to be cautious. Survey protocols were adjusted to avoid this after that point. This decision was filed in a PAP amendment before the second wave of data collection.

and Appendix Table A5 reports results with the problematic surveys included. Results do not substantially change, although confidence sets are wider. Appendix Table A7 demonstrates that results also hold if beliefs about the probability that a helmet will save one's life are winsorized at the 2nd and 98th percentiles.

Recall from section 3 that VSL is a function of one's expected utility from being alive less one's expected utility from not being alive normalized by the marginal utility of consumption. Religious beliefs are strong in Kenya, and there is near universal belief in an afterlife. The fact that VSL falls well below annual income thus likely indicates that agents expected utility after death is high relative to their expected utility from being alive. Furthermore, I argue in section 5.7 that the coefficient of relative risk aversion implied by the difference between this estimate and an estimate orders of magnitude larger among high income Africans in Sierra Leone falls within admissible parameters, assuming that $u(\text{alive}) - u(\text{not alive})$ is the same across the populations. Hence, even though this estimate is small, it is consistent with economic theory.

5.5 The value of a statistical life year

Estimates of the value of a statistical life year in Table 6 similarly show low demand for safety. My preferred estimate of VSLY, presented in column 3 of panel a, is just USD PPP \$3.31, and the weak instrument robust confidence set excludes values above \$6.40 with 95% confidence. Across all estimates on the full sample, the largest value contained within the 95% confidence set is under \$19. Results are similar across the restricted sample of those that received information in panel b, and we may reject values above \$60 across all specifications.

5.6 Heterogeneity in the value of a statistical life

As outlined in Ashenfelter (2006), if there is large heterogeneity in VSL then estimating the parameter over a selected population may yield estimates that vary from the average VSL over populations relevant to policy decisions. The estimates presented in this paper may be interpreted as the average VSL over a broad sample of urban consumers, a sample that is directly relevant to transportation policy. However, I examine the degree to which VSL is heterogeneous across 8 different variables in Table 7 to provide insight into the extent to which the average VSL over all residents of Nairobi may differ from these estimates. I find some evidence of heterogeneity, but estimates are small relative to average VSL. Hence, it is unlikely that the VSL of an average resident of Nairobi varies from these estimates by an economically significant amount unless there is significantly more heterogeneity with respect to unobservables than observables.

In column 1, I examine heterogeneity by age. Since a younger respondent has more life years remaining in expectation, we would expect age to be negatively correlated with VSL if VSLY were homogeneous. In contrast, I estimate that the VSL of a respondent above the median age of 32 is higher by $64. This indicates that older respondents have higher VSLYs.

Hall and Jones (2007) note that VSL should theoretically increase with wealth due to concave returns to consumption in any given year of one's life. Testing this theory would require data on one's full consumption bundle. The closest value in this study is wages, which I find are not strongly related to VSL in column 2. However, this estimate does not account for the fact that wage differences may be negatively correlated with the consumption of leisure or other unobserved determinants of utility, so the result should not be interpreted as a rejection of this theory. Similarly, I see no heterogeneity on expected future wage growth as reported in column 3.

Individuals that are in better self-reported health have 60% higher estimated VSL as reported in column 4. Respondents with children have lower point estimates of VSL which could reflect higher marginal utility of money, but the difference is not statistically significant. There is no statistically significant heterogeneity by performance on a digit span recall test, years of education, or gender.

5.7 Experimental versus observational estimates

Comparisons of observational and experimental VSL estimates reported in Table 8 suggest that observational revealed preference methods do not produce consistent estimates of VSL in this sample.

Columns 1 - 4 report estimates of VSL obtained from the demand model that uses empirical risk estimates as a proxy for beliefs. Estimates tend to fall outside of experimental confidence sets, they are imprecisely estimated, and they vary substantially based on the set of controls and fixed effects included. The largest estimate is about \$14,940 without controls in column 1, and the smallest is below -\$4,000 with controls selected via double-post LASSO in column 3. Column 4 includes taxi terminal fixed effects. Since the set of possible drivers and routes is small in this case, we would expect private information about risk to be minimal and thus the empirical estimate of mortality risk may be more effective. In fact, the point estimate of \$994 falls in some experimental confidence intervals. However, this value is imprecisely estimated (the 95% confidence interval spans from under -\$19,000 to over \$20,000) and the estimate is sensitive to the set of covariates included. The fact that the estimate falls within a plausible range based on the experimental estimates thus seems largely a product of chance, although I cannot reject the null hypothesis that it is consistent.

I also find that estimating $VSL_i = \frac{v_i}{\Delta r_i^*}$ yields biased estimates. The mean VSL estimate obtained from this approach is over $370,000 as reported in column 5, and the confidence interval excludes values below $347,000. Hence, one may easily reject the null hypothesis that the estimates match experimental estimates of average VSL (the preferred estimate is recreated in column 6 for comparison).

In theory, heterogeneity in VSL across the sample could cause observational and experimental estimates to vary because they weight observations differently. Results from section 5.6 indicate that heterogeneity is modest in this sample and thus unlikely to explain the differences. To further rule out this explanation, I calculate a weighted experimental estimate in column 7 which weights observations equally to the estimate from column 3 in expectation.[22] The point estimate is about $434 and the weak IV robust confidence set spans from $77 to $989. The fact that this value exceeds $420 is consistent with the fact that VSL is heterogeneous in this sample. However, the magnitude of heterogeneity cannot rationalize the observational estimates, and this estimate is in fact further from the point estimate in column 3 than the unweighted estimate.

The fact that observational estimates do not match experimental estimates is consistent with the findings presented in section 5.2 that elicited beliefs are not correlated with determinants of empirical risk. Observational estimates use empirical risk as a proxy for belief, but the belief data indicates that the two variables are poorly correlated in this setting because agents form their views based on information that is unlikely to yield unbiased assessments.

[22]OLS weights observations proportionally to the square of empirical risk reduction after partialling out controls, $(\Delta\dot{r}_i^*)^2$. Observations in the non-interacted IV are weighted equally in expectation. Hence, I assign weights $w_i = (\Delta\dot{r}_i^*)^{-2}$.

5.8 Comparisons to existing revealed preference VSL estimates

The VSL estimates presented in this paper fall below most existing revealed preference estimates from low-income settings. However, values are similar, up to an order of magnitude, to estimates calculated over populations with similar income levels. Figure 4 plots point estimates and standard errors from all revealed preference estimates of VSL from low-income settings that I am aware of.[23] This includes Kremer et al. (2011), León and Miguel (2017), Berry et al. (2020), Ito and Zhang (2020), and Shrestha (2020). In addition, I plot a revealed preference estimate from Greenberg et al. (2021) which examines a population of US soldiers for comparison to a high-income country.

The dispersion of VSL estimates is large across existing studies. For instance, León and Miguel (2017) estimate an average VSL over \$700,000 among wealthy international travelers, with an upper confidence interval over \$1.4 million. In contrast, one of the median VSL estimates presented in Berry et al. (2020) is negative, and the confidence interval reported in Shrestha (2020) includes values below -\$600,000. To accommodate this range of values, I plot estimates using a modified scale that is approximately linear near zero but logarithmic away from zero. In addition, I truncate values below -\$1 in panel b.

Point estimates presented in this paper are lower than most previous estimates, with the exception being the negative VSL estimated in Berry et al. (2020). However, it is a similar order of magnitude to estimates among populations with similar income levels. For instance, Kremer et al. (2011) estimate a value of about USD \$1,000 to averting a child death in rural Kenya, and Berry et al. (2020) estimate a VSL below \$0 and just over \$4,000 in Ghana. Estimates orders of magnitude larger than those presented in this paper were calculated over much wealthier populations.

[23] I convert estimates to 2022 dollars using the change in the CPI from January of the paper's publication year to November, 2022.

For instance, Ito and Zhang (2020) consider a population with an average income close to \$10,000 in 2022 terms and estimate a VSL of about \$40,000. And León and Miguel (2017) study a population of wealthy international travelers with a median income of about \$75,000 in 2022 terms. Furthermore, priors studies have lacked variation to separately identify the value of mortality risk reductions from the value of non-fatal risk reduction. By updating beliefs only about the mortality protection of helmets, this study identifies the value of mortality risk reduction alone, so one would expect it to be smaller than prior estimates.

If we take all VSL estimates as given, there is non-linear growth of VSL with income consistent with Hall and Jones (2007). As a thought experiment, one may assume that one's utility from being alive is constant across income levels and then calculate what curvature of marginal utility of consumption would be required to rationalize the differences in VSL observed across low and high income settings. Assuming a CRRA utility function, the differences between my preferred point estimate of VSL and those reported in León and Miguel (2017) would imply a coefficient of relative risk aversion of $\theta < 3$, and $\theta < 2$ if I instead consider a VSL estimate of \$3,000 from the upper range of confidence intervals in this paper.[24] These values fall within admissible estimates of θ, suggesting that the estimates of VSL presented in this paper are theoretically consistent with values orders of magnitude larger estimated over high-income populations.

5.9 Policy implications of experimental VSL estimates

The welfare implications of the experimental VSL estimates presented in this paper depend on the values that are currently used in benefit-cost analysis in sub-Saharan Africa. If decision makers employ values similar to this paper, such as Kremer et al. (2011), then benefit-cost ratios will be

[24] $\theta \approx \frac{\log(700{,}000/220)}{\log(75{,}000/4{,}750)}$

similar to those implied by this paper. But if larger VSL estimates are considered, then switching to the estimates from this paper could yield large welfare gains by better aligning policy with preferences.

I selected 5 recently published benefit-cost analyses in Kenya to examine how these VSL estimates affect policy conclusions. I identified relevant studies by searching for benefit-cost analyses on Google Scholar that referenced Kenya and utilized VSL. The 5 analyses selected, which come from 4 different studies, are the 5 most recent that I was able to locate. Two benefit-cost analyses are from Babagoli et al. (2022) which evaluates a program to provide menstrual cups and sanitary pads to young women, one is from Mwai et al. (2023) which examines primary health care investment, one is from Hamze et al. (2017) examining a cleft lip and palate repair program, and the final value was published in Oyugi et al. (2023) which studies a free maternity program.[25] I examine how my VSL estimates change the policy conclusions of these studies by replacing the original VSL and VSLY estimates used to value health benefits with the preferred estimates from this paper.

Figure 5 shows that benefit-cost ratios (BCRs) drop substantially when I replace the studies' VSL and VSLY values with the estimates from this paper. BCRs fall by over 99% on average, and in 4 out of 5 instances the ratio falls from a value above 1 to a value below 1, changing the conclusion about whether the programs are welfare maximizing. These differences are driven by the fact that the studies value lives and life years saved an order of magnitude higher than consumers in this study: the studies value a disability-adjusted life year saved at over $3,000 or a statistical life saved at over $200,000. Hence, this paper suggests that prioritizing more programs

[25] If multiple estimates were published for the same program, I selected the one that produces the smallest benefit-cost ratio.

aimed at improving consumption and income levels over health interventions could be welfare enhancing, reflecting the high marginal value of consumption.

5.10 Potential threats to the validity of VSL estimates

Evidence suggests that the experimental estimates of the value of a statistical life presented in this paper are robust to a large set of potential confounds. First, it is unlikely that experimenter demand effects confound because respondents did not report updated beliefs in response to all of the information they were presented with. If demand effects were present, we would expect respondents to report changes in beliefs in response to information about motorcycle risks and helmet effectiveness, but there were no changes in posteriors about the risks of motorcycles. Furthermore, there were only changes in beliefs about helmet effectiveness after surveyors learned to respond to detailed questions about the information they were presenting and respondents rarely reported that their posterior equaled a study estimate exactly.

Second, cognitive frictions in understanding small probabilities could cause respondents to systematically misreport beliefs. The study was designed to limit this risk by focusing on one of the largest sources of mortality risk that is faced by a broad segment of the population. Furthermore, respondents that responded with risks well above empirical estimates provided sensible reasons for departing from the data, and we found that stated beliefs were broadly consistent across different measures. These results suggest that the survey elicited a meaningful signal about beliefs, particularly since treatment assignment is used as an instrument to create robustness against measurement error.

The similarity of estimates based on the interacted and non-interacted version of the IV also provides evidence against this confound. The treatment only IV weights observations proportionally to random treatment assignment, whereas the interacted version weights observations proportionally to their prior about dying in a motorcycle accident squared. If respondents that communicated high risks struggled to communicate small probabilities but truly believed that the risks were lower, then we would expect VSL estimates to be higher when these observations receive less weight in the treatment only IV. However, estimates are similar across the two sets of IVs, and a Hansen J test fails to reject the equality of the estimates ($p > 0.5$). In addition, Table 7 shows that effects are not heterogeneous with respect to education or performance on a digit span recall test, which one would expect to be positively correlated with VSL if this confound were binding.

Third, inaccurate measures of willingness to pay could bias estimates. One possibility is capital constraints. The experiment aimed to mitigate this risk relative to existing work by using a willingness to accept instead of a willingness to pay mechanism, meaning respondents received either a free helmet or a cash payment rather than paying for a helmet. This ensures that liquidity constraints do not bind, but it is possible that capital market frictions could still inflate the value of cash. This is unlikely to significantly bias results since average willingness to pay is small relative to incomes. A second possible avenue of bias is from the Becker et al. (1964) mechanism. A literature suggests that different versions of the mechanism, such as willingness to pay versus willingness to accept, can produce different valuations (Isoni et al., 2011). But these differences are unlikely to bias VSL estimates since the paper leverages changes in WTP with respect to changes in beliefs and not levels, and Cole et al. (2020) and Berry et al. (2020) demonstrate that BDM mechanisms generate variation similar to take it or leave it offers among similar populations. Moreover, the survey implemented a practice round of the game to limit confusion, and surveyors reported strong understanding of the game among respondents.

Fourth, consumers may change future ridership in response to information or receiving a helmet. Among those that received cash, we see no change in planned future ridership. Those that received a helmet do report that they plan to take about one additional trip per week on average, but there are not significant differences across treatment arms. For robustness, Appendix Table A6 panel a reports VSL estimates in which the risk reduction offered by a helmet is recalculated using planned future ridership. Estimates are similar to Table 5.

Finally, I assume a helmet lifespan of 5 years, which is recommended by the manufacturer and was presented in the survey. But respondents may expect to use the helmet for more or less time. Respondents reported planning to use the helmet for 5.4 years on average, with a median of 5, suggesting that this is not likely an important source of bias. Consistent with this view, Appendix Table A6 panel b shows that VSL estimates are similar if we recompute risk reductions using perceived helmet lifespans.

6 Public policy implications of helmet demand

The primary goal of this study is to generate experimental estimates of VSL and examine how they compare to observational estimates. However, this study is among the first to rigorously estimate demand for motorcycle helmets in East Africa. This is important since helmet use is low and road accidents are a leading, and rapidly rising, cause of death among young adults. Discussions with helmet producers and road safety organizations attributed low helmet use to a lack of demand prior to this study, not supply side frictions. Although this study is not designed to evaluate programs to improve helmet use, it tests whether demand is in fact low at observed prices. In addition, the study provides insight into the extent to which information about the risks of motorcycle taxis, the effectiveness of helmets at preventing death, or subsidies could affect demand.

Table 9 indicates that helmet demand is not low because consumers do not understand the safety benefits of helmets in this setting. The effects of the treatment are small and average, and presenting consumers with information reduces their willingness to pay for helmets. Although the average effect is small, there is heterogeneity based on beliefs about the dangers of a helmet. Column 6 indicates that an agent with baseline mortality beliefs in the 90th percentile values a helmet about $5 less on average after being presented with information that helmets are only 42% effective compared to a consumer that receives no information or is exposed to information that helmets are 70% effective.

Although use of helmets is rare in Nairobi, there is evidence of unmet demand for the helmets offered in the study. Figure 6 plots inverse demand and the elasticity of demand among control respondents. [26] The wholesale price of helmets is denoted by a vertical red line. I estimate that over 75% of motorcycle taxi passengers are willing to pay at least the wholesale cost of helmets, and about 60% are willing to pay Ksh 1,000 or more, which is over 170% of the wholesale cost.

High demand is reconciled with low VSL estimates by two facts. First, as outlined in section 5.1, consumers view motorcycles as extremely dangerous. So the change in willingness to pay induced by the information interventions is small relative to the change in mortality risk. Second, consumers value characteristics of helmets other than protection against fatal accidents. I estimate that the average willingness to pay for a helmet with zero mortality risk reduction would be about $28 in this sample ($p < .01$).[27] This study was not designed to rigorously identify determinants of helmet demand other than mortality risk protection. However, respondents qualitatively reported valuing protection against non-fatal injuries, which could result in large financial losses due to foregone wages and medical bills.

[26] The elasticity of demand is estimated using a local polynomial regression adapted from Berry et al. (2020).

[27] This is based on a two-stage least squares regression of willingness to pay on an intercept and mortality risk reduction instrumented for via treatment assignment.

Further supporting the view of unmet demand, the helmet manufacturer sent vans to several motorcycle taxi stands in Nairobi after seeing these results and found that they were able to quickly sell out a stock of helmets. This suggests that the retail market for helmets may be thin. Qualitative evidence supports this conclusion. About half of respondents said that they had not previously purchased a helmet because they had never thought to purchase one, typically because they had never seen one in a store before or believed that they were sold bundled with motorcycles. We were able to locate 16 stores selling helmets in the Nairobi area. The stores generally focused on selling motorcycles with no advertising for helmets and surveyors reported that they were in areas that would not often be visited by typical consumers. No members of the field team had seen a motorcycle helmet offered in an enterprise that they visited prior to working on the study. The low retail availability of helmets may be driven in part by the fact that the manufacturer used for this study had only been producing helmets for about a year and had not engaged in significant advertising efforts at the time we conducted the study. We are not aware of any other high quality and affordable helmets in the market, so the helmets used in the study were a relatively new product. This evidence is only suggestive since the study was not designed to rigorously test for supply-side frictions. However, the demand estimates and qualitative results pointing to an under-provision of high-quality motorcycle helmets indicates that research studying the retail market could be effective at increasing helmet use.

7 Conclusion

This study presents among the first experimental estimates of the value of a statistical life. I estimate that urban East African consumers have low demand for safety. These estimates are much lower than those typically used to inform decisions in low-income settings, suggesting that there is room to substantially improve welfare by better aligning the decisions of governments and NGOs with individuals' preferences. For instance, GiveWell, a large non-profit that matches donor funds to charities, weights averting a death of someone aged 20-24 118 times as highly as doubling their consumption for a year.[28] These weights are based in part on VSL estimates of over USD $38,000 in a similar context (Redfern et al., 2019). In contrast, the VSL estimates presented in this paper indicate that consumers would value a doubling of their annual income more than 10 times as highly as averting a death on average. This suggests that agents could be made better off by funding programs targeted at improving incomes versus safety in many instances. Consistent with this view, I find that the conclusions of 4 out of 5 benefit-cost analyses that I examined flipped when the experimental VSL estimates from this paper were used in place of the originally used values.

I caution that these estimates of VSL are valid only in instances in which the opportunity cost of a safety program is consumption among Kenyans. This trade-off is applicable when aid organizations decide between programs aimed at improving incomes versus health or for East African governments choosing whether to fund health and safety programs. But it is not relevant when the financial cost of a program is paid for by high-income consumers since in these cases the welfare costs of the program are a function of the high-income, not low-income, consumer's marginal utility of consumption. For instance, if a European government were conducting a benefit-cost analysis of a program aimed at reducing an externality causing excess mortality in East Africa, then European and not East African consumers would bear the costs of the program. The correct VSL in this case would be a function of an East African's average utility of averting a death normalized by a European consumer's marginal utility of consumption. This value cannot be iden-

[28]Based on the 2020 update on GiveWell's moral weights.

tified without additional assumptions since utils are subjective (M and Robbins, 1934). But since economic theory indicates that the marginal utility of a dollar is strongly related to income levels, a reasonable rule of thumb would be to price lives saved by a European's VSL.

In addition to producing experimental estimates of VSL, this paper is the first to compare experimental and observational estimates. This provides some of the first evidence about the extent to which biased mortality beliefs and endogeneity, potential problems in observational VSL estimates that have been identified in studies such as Ashenfelter (2006), affect VSL estimates in practice. I find that cross-sectional estimates of VSL are biased, and I present evidence that empirical estimates of mortality risk align poorly with beliefs because agents do not learn from representative data. Barriers to objective information about risk – such as low internet access, low access to high quality journalism, and weak government institutions – raise the concern that agents in other low-income settings are also likely to have biased beliefs. This suggests the need for future research to examine whether beliefs about risk are accurate in other settings. If not, then estimates of VSL that use elicited beliefs rather than empirical risk estimates are important to guide optimal trade-offs between safety and risk.

References

Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2022, 12). When Should You Adjust Standard Errors for Clustering? *The Quarterly Journal of Economics 138*(1), 1–35.

Anderson, T. W. and H. Rubin (1949). Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations. *Source: The Annals of Mathematical Statistics 20*(1), 46–63.

Andrews, D. W., M. J. Moreira, and J. H. Stock (2007, 7). Performance of conditional Wald tests in IV regression with weak instruments. *Journal of Econometrics 139*(1), 116–132.

Ashenfelter, O. (2006, 3). Measuring the Value of a Statistical Life: Problems and Prospects*. *The Economic Journal 116*(510), C10–C23.

Ashenfelter, O. and M. Greenstone (2004, 5). Estimating the Value of a Statistical Life: The Importance of Omitted Variables and Publication Bias. *American Economic Review 94*(2), 454–460.

Ashenfelter, O. and M. Greenstone (2015, 7). Using Mandated Speed Limits to Measure the Value of a Statistical Life. *https://doi.org/10.1086/379932 112*(1).

Attanasio, O. P. and K. M. Kaufmann (2017, 8). Education choices and returns on the labor and marriage markets: Evidence from data on subjective expectations. *Journal of Economic Behavior & Organization 140*, 35–55.

Babagoli, M. A., A. Benshaul-Tolonen, G. Zulaika, E. Nyothach, C. Oduor, D. Obor, L. Mason, E. Kerubo, I. Ngere, K. F. Laserson, R. T. Edwards, and P. A. Phillips-Howard (2022, 9). Cost-Effectiveness and Cost–Benefit Analyses of Providing Menstrual Cups and Sanitary Pads to Schoolgirls in Rural Kenya. *Women's Health Reports 3*(1), 773.

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.

Banzhaf, H. S. (2014, 9). Retrospectives: The Cold-War Origins of the Value of Statistical Life. *Journal of Economic Perspectives 28*(4), 213–26.

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.

Berry, J., G. Fischer, and R. Guiteras (2020, 4). Eliciting and Utilizing Willingness to Pay: Evidence from Field Trials in Northern Ghana. *https://doi.org/10.1086/705374 128*(4), 1436–1473.

Breza, E., S. Kaur, and Y. Shamdasani (2021, 10). Labor Rationing. *American Economic Review 111*(10), 3184–3224.

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

Cole, S., A. N. Fernando, D. Stein, and J. Tobacman (2020, 4). Field comparisons of incentive-compatible preference elicitation techniques. *Journal of Economic Behavior & Organization 172*, 33–56.

Delavande, A. (2008, 8). PILL, PATCH, OR SHOT? SUBJECTIVE EXPECTATIONS AND BIRTH CONTROL CHOICE*. *International Economic Review 49*(3), 999–1042.

Delavande, A. (2014, 8). Probabilistic Expectations in Developing Countries. *http://dx.doi.org/10.1146/annurev-economics-072413-105148 6*, 1–20.

Finlay, K., L. Magnusson, and M. E. Schaffer (2016, 10). WEAKIV: Stata module to perform weak-instrument-robust tests and confidence intervals for instrumental-variable (IV) estimation of linear, probit and tobit models. *Statistical Software Components*.

Greenberg, K., M. Greenstone, S. P. Ryan, and M. Yankovich (2021). The Heterogeneous Value of a Statistical Life: Evidence from U.S. Army Reenlistment Decisions. *National Bureau of Economic Research Working Paper Series No. 29104*.

Greenstone, M. and B. K. Jack (2015, 3). Envirodevonomics: A Research Agenda for an Emerging Field. *Journal of Economic Literature 53*(1), 5–42.

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.

Habyarimana, J. and W. Jack (2015, 8). Results of a large-scale randomized behavior change intervention on road safety in Kenya. *Proceedings of the National Academy of Sciences of the United States of America 112*(34), E4661–E4670.

Hall, R. E. and C. I. Jones (2007, 2). The Value of Life and the Rise in Health Spending. *The Quarterly Journal of Economics 122*(1), 39–72.

Hamze, H., A. Mengiste, and J. Carter (2017). The impact and cost-effectiveness of the Amref Health Africa-Smile Train Cleft Lip and Palate Surgical Repair Programme in Eastern and Central Africa. *The Pan African medical journal 28*, 35.

Isoni, A., G. Loomes, and R. Sugden (2011). The Willingness to Pay—Willingness to Accept Gap, the "Endowment Effect," Subject Misconceptions, and Experimental Procedures for Eliciting Valuations: Comment. *American Economic Review 101*.

Ito, K. and S. Zhang (2020, 5). Willingness to pay for clean air: Evidence from air purifier markets in China. *Journal of Political Economy 128*(5), 1627–1672.

Kremer, M., J. Leino, E. Miguel, and A. P. Zwane (2011, 2). Spring Cleaning: Rural Water Impacts, Valuation, and Property Rights Institutions*. *The Quarterly Journal of Economics 126*(1), 145–205.

Lee, J. M. and L. O. Taylor (2019, 11). Randomized Safety Inspections and Risk Exposure on the Job: Quasi-experimental Estimates of the Value of a Statistical Life. *American Economic Journal: Economic Policy 11*(4), 350–74.

León, G. and E. Miguel (2017, 1). Risky Transportation Choices and the Value of a Statistical Life. *American Economic Journal: Applied Economics 9*(1), 202–28.

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).

M, H. W. and L. Robbins (1934). An Essay on the Nature and Significance of Economic Science. *Journal of the Royal Statistical Society 97*(2).

McKenzie, D., J. Gibson, and S. Stillman (2013, 5). A land of milk and honey with streets paved with gold: Do emigrants have over-optimistic expectations about incomes abroad? *Journal of Development Economics 102*, 116–127.

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

Mwai, D., S. Hussein, A. Olago, M. Kimani, D. Njuguna, R. Njiraini, E. Wangia, E. Olwanda, L. Mwaura, and W. Rotich (2023, 3). Investment case for primary health care in low- and middle-income countries: A case study of Kenya. *PLOS ONE 18*(3), e0283156.

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.

Oyugi, B., O. Nizalova, S. Kendall, and S. Peckham (2023, 2). Does a free maternity policy in Kenya work? Impact and cost–benefit consideration based on demographic health survey data. *The European Journal of Health Economics*.

Rabin, M. (2002). Inference by Believers in the Law of Small Numbers. *Source: The Quarterly Journal of Economics 117*(3), 775–816.

Redfern, A., M. Gould, M. Chege, S. Li, F. A. Garay, and W. Slotznick (2019). Beneficiary preferences: Findings from Kenya and Ghana. Technical report, IDinsight.

Rohlfs, C., R. Sullivan, and T. Kniesner (2015). New Estimates of the Value of a Statistical Life Using Air Bag Regulations as a Quasi-experiment. *American Economic Journal: Economic Policy 7*(1), 331–59.

Shrestha, M. (2020, 2). Get Rich or Die Tryin': Perceived Earnings, Perceived Mortality Rates, and Migration Decisions of Potential Work Migrants from Nepal. *The World Bank Economic Review 34*(1), 1–27.

Figures

Figure 1: Survey locations

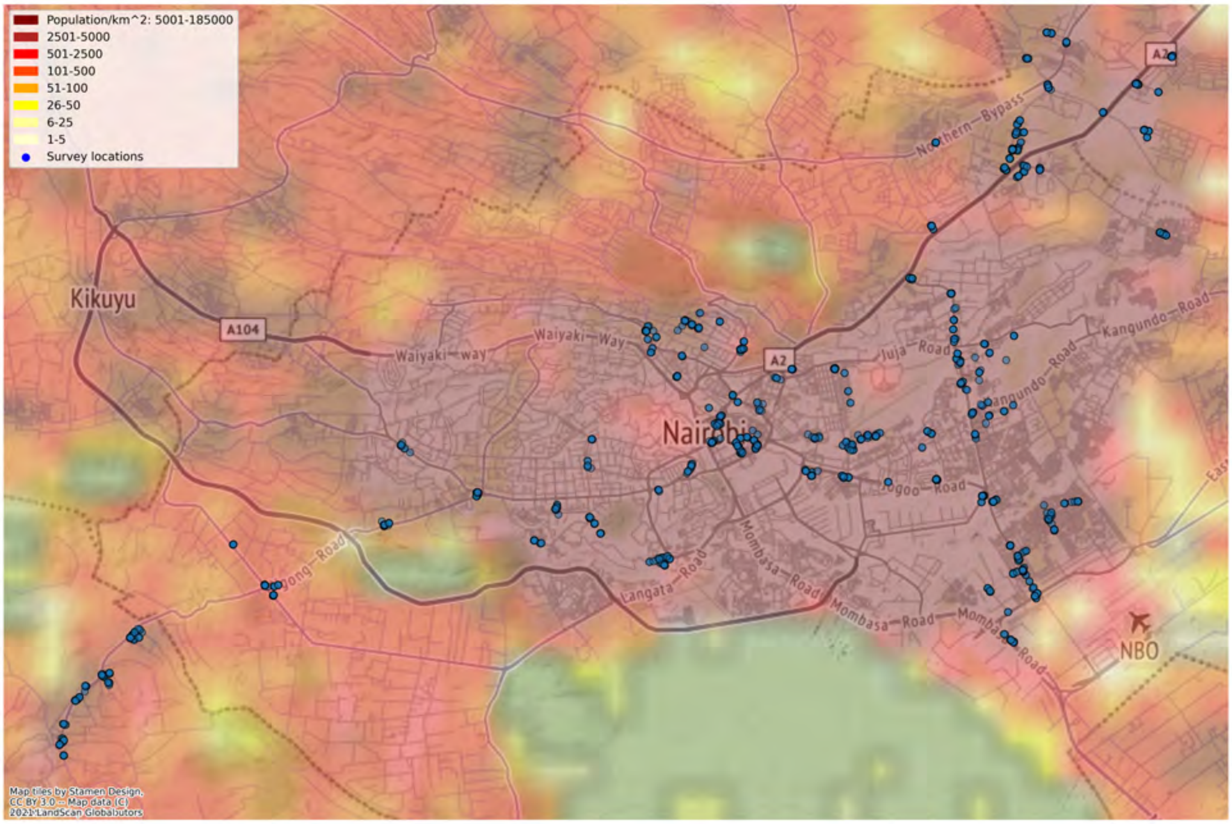
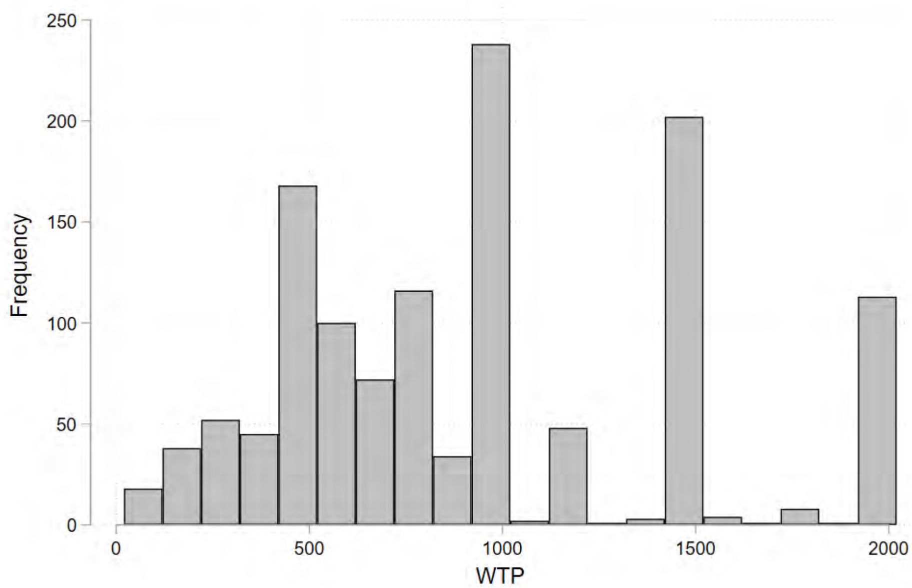


Figure 1 plots survey locations over a map of Nairobi. Color denotes population, where no color denotes no residents. Map data is from Open Street Map. Population data is from the LandScan Global 2021 data set produced by Oak Ridge National Laboratory.

Figure 2: Distribution of helmet bids (Kenyan shillings)

(a) A. Histogram of bids, restricted axis



(b) B. Histogram of bids, full distribution

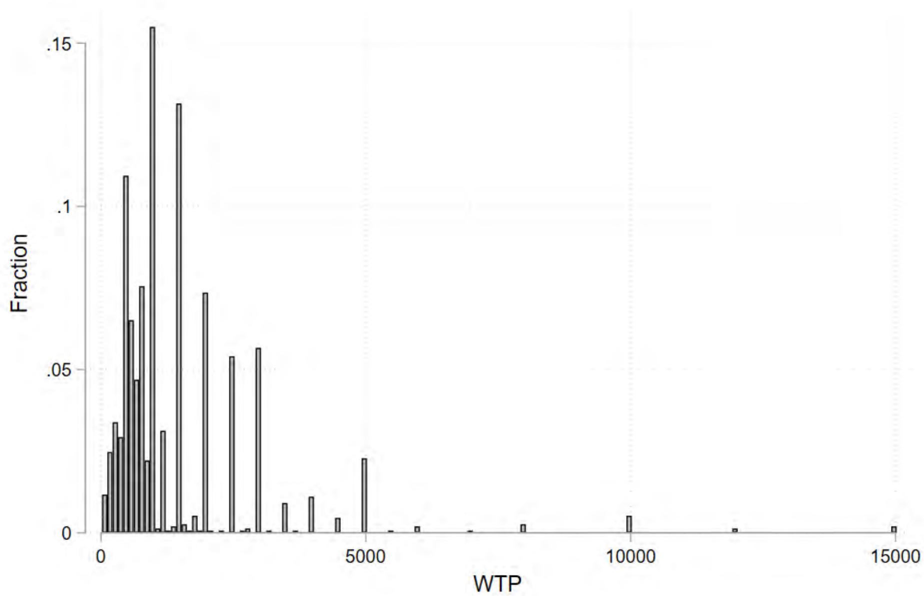


Figure 2 plots the distribution of willingness to pay for helmets in Kenyan shillings (Ksh). Panel A reports a histogram of bids, excluding outliers above Ksh 2,000 for clarity. Panel B reports the same histogram across the full distribution of bids.

Figure 3: Information sources used to form beliefs about motorcycle safety

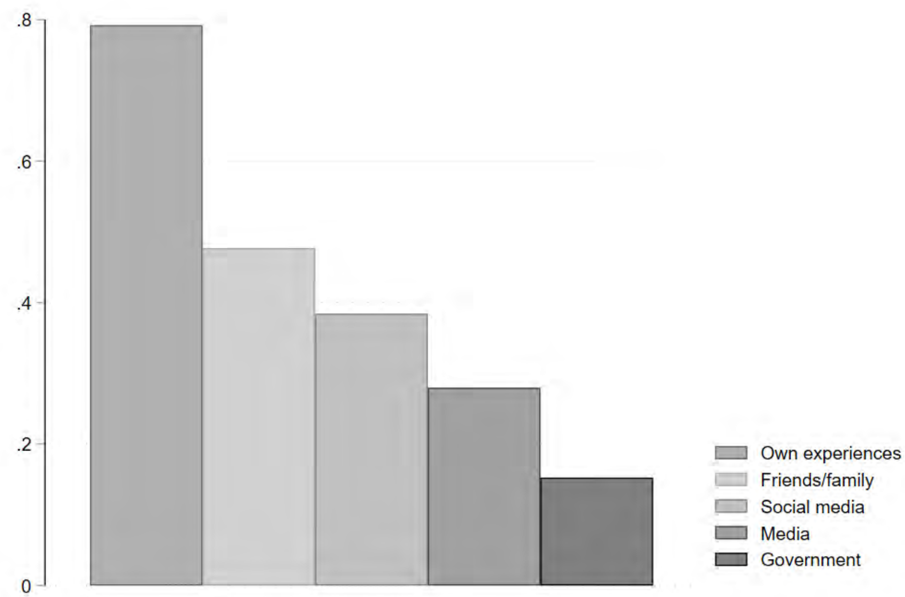


Figure 3 plots the information sources that respondents reported using to form beliefs. Respondents were able to select multiple options, so the columns do not add to 1.

Figure 4: VSL estimates across studies

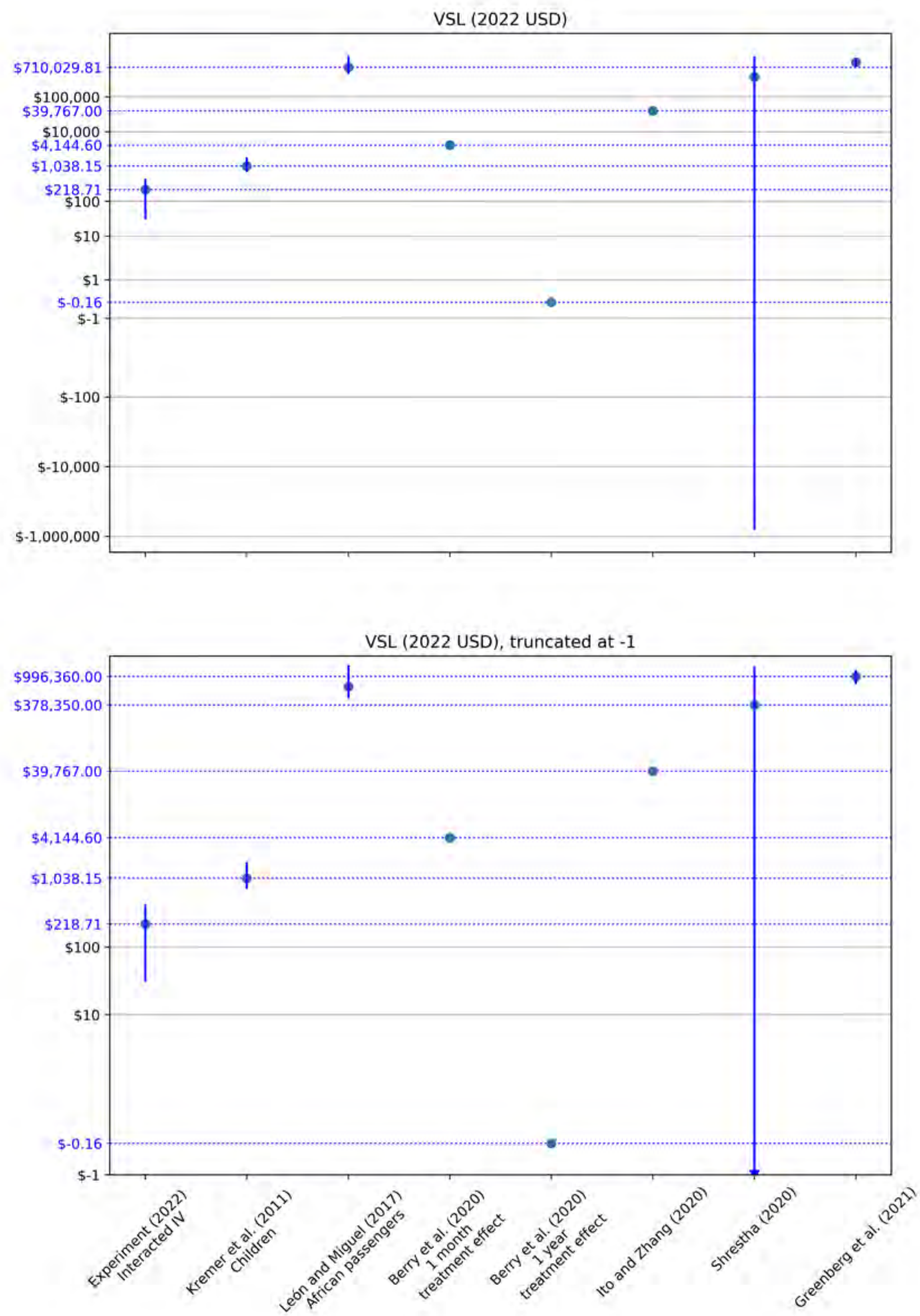


Figure 4 plots VSL estimates from this paper, Kremer et al. (2011), León and Miguel (2017), Berry et al. (2020), and Ito and Zhang (2020). All estimates are presented in 2022 USD calculated by inflating based on the paper's publication year using the CPI inflation calculator. Panel a presents estimates in dollar terms. Panel b reports VSL relative to annual income for papers in which income was presented.

Figure 5: Effect of study VSL on published benefit-cost ratios

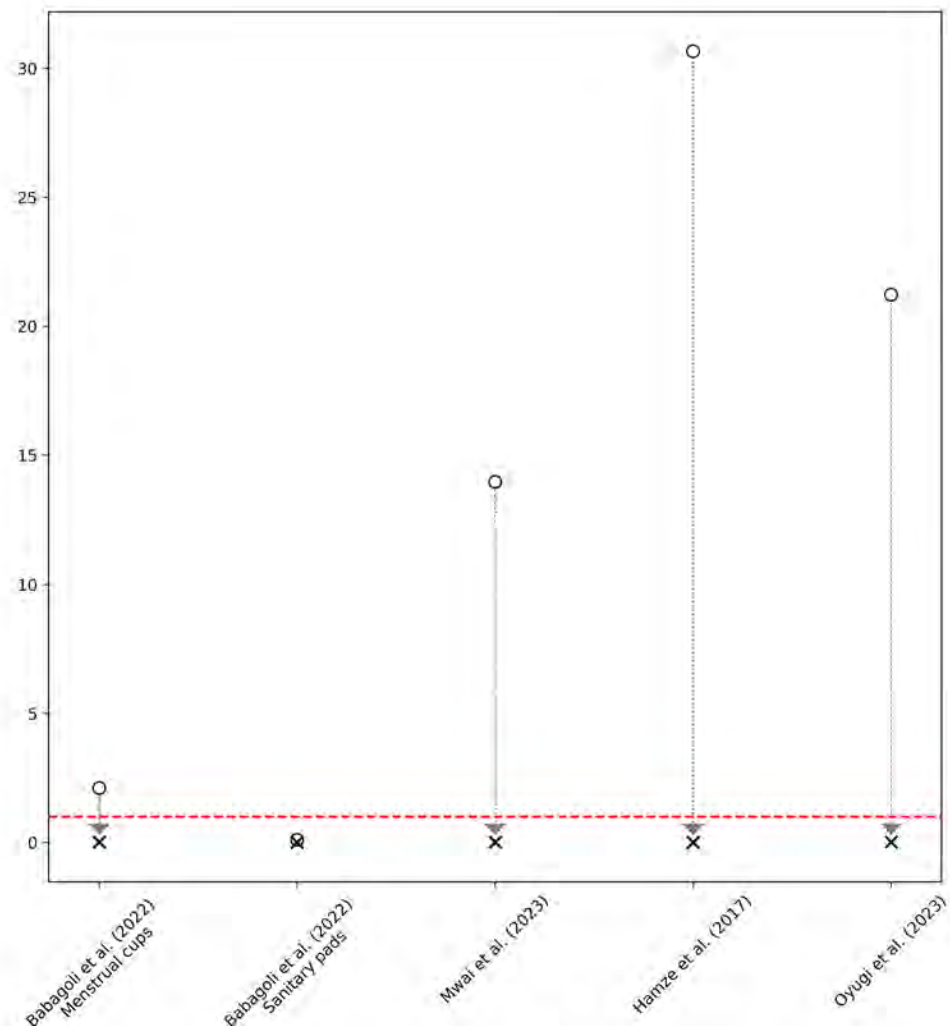
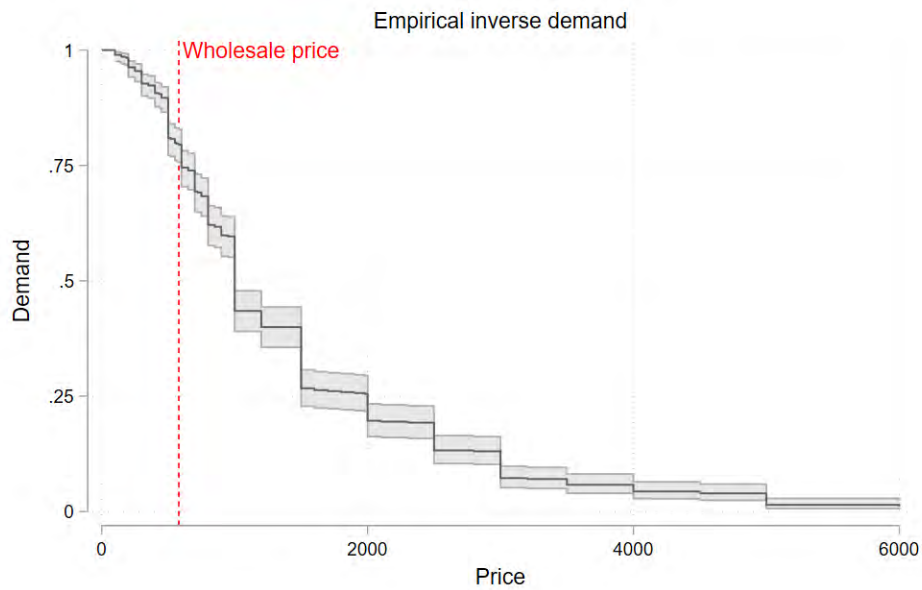


Figure 5 examines how the benefit-cost ratios (BCRs) of published benefit-cost analyses of Kenyan programs change when the study's original VSL or VSLY is replaced with the preferred estimates from this study. Hollow circles denote the original BCR estimate, and an x denotes the revised estimate. The first two estimates are from Babagoli et al. (2022). The third estimate is from Mwai et al. (2023), the fourth is from Hamze et al. (2017), and the final figure is from Oyugi et al. (2023). The horizontal red line is at $BCR = 1$, the threshold for benefits exceeding costs.

Figure 6: Demand for helmets

(a) Inverse demand, control group



(b) Elasticity of demand, control group

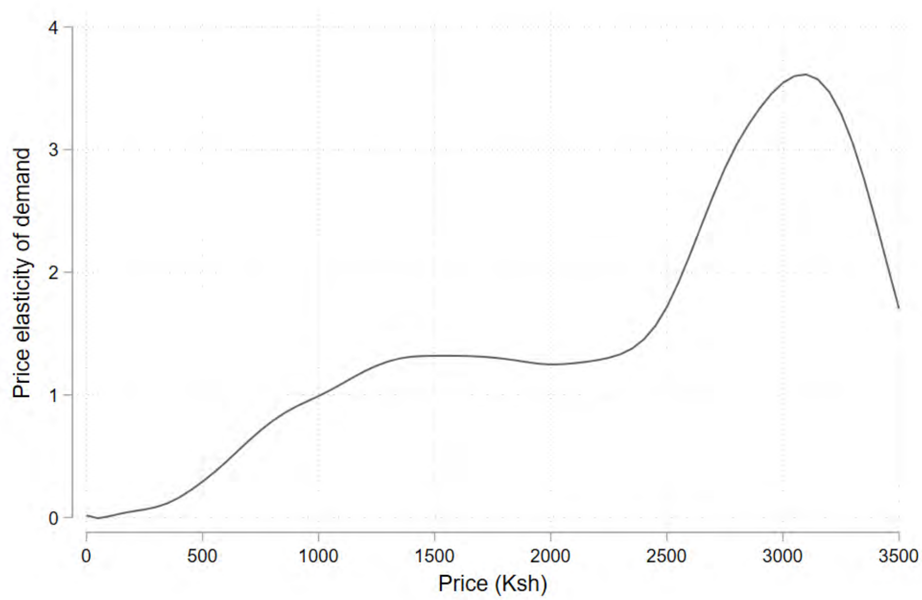


Figure 6 plots estimates of demand for helmets among observations in the control and pure control arms. The plot of demand includes a pointwise confidence interval. The plot of the elasticity of demand is based on a local polynomial estimation adapted from Berry et al. (2020). The vertical line denotes the wholesale price of helmets, which was Ksh 580 during the study. The figures exclude 35 observations for which surveyors reported motorcycle taxi drivers pretending to be passengers.

Tables

Table 1: Summary statistics and balance: Demographics

	(1) Control Summary statistics	(2) Pure control - control	(3) Low treatment - Control	(4) High treatment - Control	(5) High - low treatment
Age	32.735	0.105	0.275	0.667	0.392
	[8.126]	(1.159)	(0.572)	(0.589)	(0.590)
Female	0.352	-0.040	-0.064**	-0.049	0.014
	[0.478]	(0.060)	(0.030)	(0.030)	(0.029)
Health (1-5)	3.403	-0.034	-0.058	-0.009	0.049
	[0.667]	(0.086)	(0.043)	(0.045)	(0.043)
Life expectancy	81.438	-0.134	0.462	-0.124	-0.586
	[6.947]	(0.840)	(0.427)	(0.439)	(0.413)
Employed	0.898	-0.051	-0.041*	-0.046**	-0.005
	[0.303]	(0.043)	(0.022)	(0.022)	(0.022)
Income (PPP, ’000s USD)	6.730	-0.147	0.625	1.012	0.386
	[8.260]	(1.443)	(0.817)	(0.845)	(0.936)
$\mathbb{E}$[Wage in 5 years/Wage today]	6.166	-0.636	-0.907	-1.353*	-0.446
	[11.685]	(1.761)	(0.719)	(0.743)	(0.674)
1(children)	0.774	0.033	-0.002	-0.013	-0.011
	[0.418]	(0.053)	(0.027)	(0.028)	(0.027)
Digit span recall	3.020	-0.156	-0.005	-0.011	-0.007
	[1.391]	(0.183)	(0.089)	(0.091)	(0.088)
Years of education	12.111	-0.468	0.162	-0.026	-0.188
	[2.869]	(0.371)	(0.184)	(0.189)	(0.181)
1(primary school complete)	0.964	-0.035	0.007	-0.022	-0.029**
	[0.186]	(0.026)	(0.013)	(0.013)	(0.013)
1(secondary school complete)	0.722	-0.047	0.024	-0.008	-0.032
	[0.449]	(0.058)	(0.029)	(0.029)	(0.028)
1(college degree)	0.241	-0.054	-0.003	0.014	0.018
	[0.428]	(0.055)	(0.028)	(0.028)	(0.027)
N (exc. control in col. 2-4)	452	80	531	473	
p-value of joint orthogonality		0.743	0.173	0.204	0.273

Standard deviations in brackets. Standard errors in parenthesis.

Column (1) reports the mean and standard deviation of the indicated variable across the control group. Columns (2) - (5) report the difference in mean of each variable across treatment arms, where terms in parentheses are standard errors of this pairwise t-test. Missing values are imputed using the median across the full sample for the joint test.

Table 2: Summary statistics and balance: Motorcycle taxi use

	(1) Control Summary statistics	(2) Pure control - control	(3) Low treatment - Control	(4) High treatment - Control	(5) High - low treatment
Trips/week	7.531	1.178	0.192	-0.375	-0.567*
	[5.242]	(0.725)	(0.348)	(0.358)	(0.343)
Average trip length (minutes)	19.593	0.334	-0.725	-1.276*	-0.551
	[12.567]	(1.538)	(0.726)	(0.746)	(0.672)
$\mathbb{E}$[Deaths/10,000 passengers, 1 year]	371.584	NA	-56.005	7.543	63.547
	[1,186.706]	NA	(78.123)	(80.394)	(78.179)
$\mathbb{E}$[Deaths/10,000 passengers, 5 years]	977.118	NA	-179.625	283.175	462.800**
	[2,479.942]	NA	(196.470)	(202.182)	(208.987)
10000*Pr(Fatal accident, 5 years)	354.133	NA	-26.685	-34.493	-7.808
	[973.710]	NA	(57.110)	(58.763)	(53.724)
Previous accident	0.488	NA	-0.011	-0.029	-0.018
	[0.500]	NA	(0.032)	(0.033)	(0.032)
Know accident victim	0.947	NA	-0.022	-0.036**	-0.013
	[0.224]	NA	(0.017)	(0.017)	(0.017)
Use motorcycle taxi: Commuting	0.781	0.027	-0.031	0.008	0.039
	[0.414]	(0.054)	(0.027)	(0.028)	(0.027)
Shopping	0.420	-0.031	-0.034	-0.017	0.018
	[0.494]	(0.064)	(0.031)	(0.032)	(0.031)
Leisure	0.261	0.057	0.074**	0.035	-0.039
	[0.440]	(0.058)	(0.029)	(0.030)	(0.029)
Deliveries	0.095	-0.012	0.018	-0.004	-0.022
	[0.294]	(0.014)	(0.019)	(0.019)	(0.019)
Emergency/hospital transportation	0.095	-0.008	-0.007	-0.011	-0.004
	[0.294]	(0.014)	(0.018)	(0.018)	(0.018)
Reason for use: Speed	0.816	0.114**	0.018	0.004	-0.014
	[0.388]	(0.045)	(0.024)	(0.025)	(0.024)
Convenience	0.717	-0.005	0.006	-0.047	-0.053*
	[0.451]	(0.061)	(0.029)	(0.030)	(0.029)
Only option	0.168	0.006	-0.025	-0.010	0.015
	[0.374]	(0.047)	(0.023)	(0.024)	(0.023)
Price	0.106	0.037	-0.008	-0.026	-0.018
	[0.308]	(0.038)	(0.019)	(0.019)	(0.018)
Safety/Avoid dangerous areas	0.069	-0.017	-0.027*	-0.009	0.018
	[0.253]	(0.033)	(0.015)	(0.015)	(0.014)
Enjoyment	0.011	-0.012	-0.002	0.008	0.010
	[0.105]	(0.015)	(0.007)	(0.007)	(0.007)
Risk information: Own experiences	0.819	NA	-0.037	-0.041	-0.004
	[0.386]	NA	(0.026)	(0.027)	(0.026)
Friends/family	0.454	NA	0.030	0.037	0.006
	[0.498]	NA	(0.032)	(0.033)	(0.032)
Social media	0.414	NA	-0.048	-0.035	0.013
	[0.493]	NA	(0.031)	(0.032)	(0.031)
Media	0.288	NA	0.016	-0.030	-0.045
	[0.453]	NA	(0.029)	(0.030)	(0.028)
Government	0.135	NA	0.016	0.026	0.010
	[0.342]	NA	(0.023)	(0.023)	(0.023)
N (exc. control in col. 2-4)	452	80	531	473	
p-value of joint orthogonality		0.008	0.116	0.180	0.157

Standard deviations in brackets. Standard errors in parenthesis.

Column (1) reports the mean and standard deviation of the indicated variable across the control group. Columns (2) - (5) report the difference in mean of each variable across treatment arms, where terms in parentheses are standard errors of this pairwise t-test. Missing values are imputed using the median across the full sample for the joint test.

Table 3: Correlates with beliefs

	(1) 10,000 × Risk	(2) 10,000 × Risk	(3) 10,000 × Risk	(4) 10,000 × Risk	(5) 10,000 × Risk	(6) 10,000 × Risk	(7) 10,000 × Risk	(8) 10,000 × Risk
Trips/week	-0.89	-0.20						
	(3.68)	((3.68))						
Trip length			-1.94	0.38				
			(1.77)	(1.58)				
Previous accident					121.28	69.77		
					(46.26)	(52.56)		
Contact in accident							256.79	192.31
							(37.63)	(55.59)
Control mean	333.22	333.22	333.22	333.22	333.22	333.22	333.22	333.22
Observations	1,427	1,427	1,427	1,427	1,427	1,427	1,427	1,427
Controls	No	LASSO	No	LASSO	No	LASSO	No	LASSO
Terminal FE	No	Yes	No	Yes	No	Yes	No	Yes

Robust standard errors in parenthesis.

Table 3 reports the correlation of demographic variables with prior beliefs about one's risk of dying in a motorcycle accident over a 5 year span. Odd numbered columns do not include any controls or fixed effects. Even numbered columns include controls include controls selected using double-post LASSO and taxi terminal/stand fixed effects (Belloni et al., 2014). All estimates include wave fixed effects and interpolate missing controls using the median of the variable.

Table 4: Effect of information on beliefs

	(1) Posterior risk	(2) Posterior risk	(3) Helmet effectiveness	(4) Helmet effectiveness	(5) Risk reduction	(6) Risk reduction, winsorized
Low treatment	-25.90	-26.44	-14.08	-14.26	-81.12	-46.41
	(21.77)	(21.97)	(0.97)	(0.99)	(18.38)	(14.12)
High treatment	6.77	4.86	-3.98	-4.36	-29.46	-0.45
	(24.91)	(25.28)	(0.88)	(0.93)	(19.70)	(16.12)
Control mean	330.97	330.97	78.68	78.68	221.79	228.62
Pr(High treatment = low treatment)	0.08	0.09	0.00	0.00	0.00	0.00
Observations	1,425	1,425	1,427	1,427	1,425	1,425
Controls	BL Risk	LASSO	BL Risk	LASSO	LASSO	LASSO
Enumerator FE	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors in parenthesis.

Columns (1) - (2) report 10,000 times the respondent's stated posterior belief about the probability that they will suffer a fatal accident over the next 5 years, the lifespan of a helmet, if they do not purchase a helmet. Columns (3) - (4) report the respondent's stated belief about the effectiveness of motorcycle helmets, expressed as the percentage reduction in mortality risk. Columns (5) - (6) report 10,000 times the likelihood that a helmet will save the respondent's life over its 5 year lifespan based on their beliefs. The outcome in column (6) is winsorized at the 2nd and 98th percentiles. All estimates include wave fixed effects. Respondents assigned to the "Low treatment" group received the results of Liu et al. (2008) which estimates that helmets are 42% effective. Those assigned to "High treatment" were presented with the results of Ouellet and Kasantikul (2006) that helmets were 70% effective in a study in Thailand. Individuals in both treatment arms received empirical estimates of their risk of a fatal accident based on ridership. All models exclude 35 observations for which surveyors reported motorcycle taxi drivers pretending to be passengers.

Table 5: Value of a statistical life: Primary estimates

Panel A: Full sample

	Raw WTP				Winsorized WTP	
	(1) Interacted	(2) Treatment only	(3) Interacted	(4) Treatment only	(5) Interacted	(6) Treatment only
VSL	210.95	340.36	218.71	302.70	96.77	245.66
	(96.43)	(254.34)	(96.12)	(257.29)	(69.73)	(189.74)
Cragg-Donald F-stat	40.12	10.99	40.59	10.59	40.59	10.59
Weak IV Robust Confidence Set	[23.62, 412.37]	[-156.99, 1,045.74]	[32.42, 419.53]	[-211.61, 1,012.75]	[-40.70, 239.36]	[-127.85, 731.72]
Inversion test	CLR	CLR	CLR	CLR	CLR	CLR
Observations	1,425	1,425	1,425	1,425	1,425	1,425
Controls	BL Risk	BL Risk	LASSO	LASSO	LASSO	LASSO
Enumerator FE	Yes	Yes	Yes	Yes	Yes	Yes

Panel B: Treated respondents only

	Raw WTP				Winsorized WTP	
	(1) Interacted	(2) Treatment only	(3) Interacted	(4) Treatment only	(5) Interacted	(6) Treatment only
VSL	388.17	759.55	384.60	782.29	173.08	407.72
	(248.25)	(468.56)	(249.80)	(487.50)	(172.95)	(324.38)
Cragg-Donald F-stat	16.84	11.24	16.35	10.39	16.35	10.39
Weak IV Robust Confidence Set	[-78.45, 1,001.51]	[-38.35, 2,356.17]	[-85.02, 1,005.81]	[-40.87, 2,523.58]	[-169.86, 577.41]	[-187.36, 1,474.13]
Inversion test	CLR	AR	CLR	AR	CLR	AR
Observations	982	982	982	982	982	982
Controls	BL Risk	BL Risk	LASSO	LASSO	LASSO	LASSO
Enumerator FE	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parenthesis.

Panel A reports regressions using observations in the control group, low treatment group, and high treatment group. Panel B uses observations from the low and high treatment arms only. Columns labeled "Interacted" use a first stage consisting of treatment indicators and treatment indicators interacted with baseline motorcycle risk beliefs as instruments for the mortality risk reduction of a helmet. Columns labeled "Treatment only" use only treatment assignment as an instrument. All models control for baseline risk beliefs. Columns (5) - (6) report estimates where willingness to pay is winsorized at the 2nd and 98th percentiles. Over-identified models report weak instrument robust confidence sets constructed using inversion of a conditional likelihood ratio (Moreira, 2003). For just-identified models, weak instrument robust confidence sets are constructed by inverting an AR test (Anderson and Rubin, 1949; Chernozhukov and Hansen, 2008). Results exclude 35 observations for which surveyors reported motorcycle taxi drivers pretending to be passengers.

Table 6: Value of a statistical life year: Primary estimates

Panel A: Full sample

	Raw WTP				Winsorized WTP	
	(1) Interacted	(2) Treatment only	(3) Interacted	(4) Treatment only	(5) Interacted	(6) Treatment only
VSLY	3.12	5.30	3.31	5.15	1.43	4.43
	(1.49)	(4.53)	(1.50)	(4.55)	(1.09)	(3.37)
Cragg-Donald F-stat	54.55	10.80	55.97	10.93	55.97	10.93
Weak IV Robust Confidence Set	[0.23, 6.20]	[-3.63, 18.18]	[0.42, 6.40]	[-3.85, 17.87]	[-0.70, 3.63]	[-2.12, 13.16]
Inversion test	CLR	CLR	CLR	CLR	CLR	CLR
Observations	1,425	1,425	1,425	1,425	1,425	1,425
Controls	BL Risk	BL Risk	LASSO	LASSO	LASSO	LASSO
Enumerator FE	Yes	Yes	Yes	Yes	Yes	Yes

Panel B: Treated respondents only

	Raw WTP				Winsorized WTP	
	(1) Interacted	(2) Treatment only	(3) Interacted	(4) Treatment only	(5) Interacted	(6) Treatment only
VSLY	5.44	16.32	5.42	15.67	2.35	8.17
	(3.93)	(10.43)	(3.89)	(9.90)	(2.72)	(6.56)
Cragg-Donald F-stat	23.75	8.78	24.32	9.48	24.32	9.48
Weak IV Robust Confidence Set	[-2.06, 14.61]	[-0.82, 59.28]	[-1.99, 14.42]	[-0.82, 53.62]	[-3.03, 8.35]	[-3.75, 31.23]
Inversion test	CLR	AR	CLR	AR	CLR	AR
Observations	982	982	982	982	982	982
Controls	BL Risk	BL Risk	LASSO	LASSO	LASSO	LASSO
Enumerator FE	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parenthesis.

Panel A reports regressions using observations in the control group, low treatment group, and high treatment group. Panel B uses observations from the low and high treatment arms only. Columns labeled “Interacted” use a first stage consisting of treatment indicators and treatment indicators interacted with baseline motorcycle risk beliefs as instruments for the mortality risk reduction of a helmet. Columns labeled “Treatment only” use only treatment assignment as an instrument. All models control for baseline risk beliefs. Columns (5) - (6) report estimates where willingness to pay is winsorized at the 2nd and 98th percentiles. Over-identified models report weak instrument robust confidence sets constructed using inversion of a conditional likelihood ratio (Moreira, 2003). For just-identified models, weak instrument robust confidence sets are constructed by inverting an AR test (Anderson and Rubin, 1949; Chernozhukov and Hansen, 2008). Results exclude 35 observations for which surveyors reported motorcycle taxi drivers pretending to be passengers.

Table 7: Heterogeneity in the value of a statistical life

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	1(Age > median)	log(wage)	1(wage growth > median)	1(health > median)	1(children)	1(digit recall score > median)	1(education > median)	1(female)
VSL	259.53	190.55	132.24	165.20	284.99	138.01	225.20	225.86
	(95.09)	(177.20)	(88.13)	(95.62)	(108.38)	(96.26)	(94.00)	(102.13)
VSL x Interaction	64.02	1.36	0.94	99.71	-72.42	24.84	-2.14	-33.40
	(28.93)	(22.83)	(30.59)	(29.55)	(41.25)	(28.66)	(31.45)	(32.56)
Cragg-Donald F-stat	18.47	19.64	21.54	18.68	18.64	18.49	19.31	16.64
Observations	1,425	1,424	1,425	1,425	1,425	1,425	1,425	1,425
Controls	LASSO	LASSO	LASSO	LASSO	LASSO	LASSO	LASSO	LASSO
Enumerator FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parenthesis.

All columns report VSL estimates across the full sample and use a first stage consisting of treatment indicators and treatment indicators interacted with baseline motorcycle risk beliefs, and each of these values interacted with the demographic variable, as instruments for the mortality risk reduction of a helmet. In columns 1, 3-4, and 5-6, the demographic variable is converted to an indicator equal to 1 if the response was above the median. The estimates exclude 35 observations for which surveyors reported motorcycle taxi drivers pretending to be passengers.

Table 8: Value of a statistical life: Comparison of non-experimental to experimental estimates

Panel A: 42% helmet effectiveness treated as truth

	Non-experimental					Experimental	
	(1) OLS	(2) LASSO	(3) LASSO	(4) LASSO	(5) $\frac{WTP_i}{\Delta r_i^*}$	(6) Preferred IV	(7) Weighted IV
VSL	14,941.50	3,026.87	-4,437.52	994.48	373,752.32	219.45	434.60
	(10,684.17)	(10,452.69)	(9,713.72)	(10,420.34)	(13,583.99)	(95.94)	(201.68)
95% CI/Confidence set	[-5,999.48, 35,882.48]	[-17460.40, 23,514.14]	[-23476.42, 14,601.38]	[-19429.38, 21,418.34]	[347,127.71, 400,376.94]	[33.54, 419.83]	[77.00, 989.48]
Cragg-Donald F-stat						40.75	12.47
Observations	1,425	1,425	1,425	1,425	1,425	1,425	1,425
Controls	None	LASSO	LASSO	LASSO	LASSO	LASSO	LASSO
Enumerator FE	No	No	Yes	No	Yes	Yes	YES
Taxi Terminal FE	No	No	No	Yes	No	No	No

Panel B: 70% helmet effectiveness treated as truth

	Non-experimental					Experimental	
	(1) OLS	(2) LASSO	(3) LASSO	(4) LASSO	(5) $\frac{WTP_i}{\Delta r_i^*}$	(6) Preferred IV	(7) Weighted IV
VSL	8,964.90	1,816.12	-2,662.51	596.69	373,752.32	219.45	434.60
	(6,410.50)	(6,271.61)	(5,828.23)	(6,252.20)	(13,583.99)	(95.94)	(201.68)
95% CI/Confidence set	[-3,599.69, 21,529.49]	[-10476.24, 14,108.48]	[-14085.85, 8,760.83]	[-11657.63, 12,851.00]	[347,127.71, 400,376.94]	[33.54, 419.83]	[77.00, 989.48]
Cragg-Donald F-stat						40.75	12.47
Observations	1,425	1,425	1,425	1,425	1,425	1,425	1,425
Controls	None	LASSO	LASSO	LASSO	LASSO	LASSO	LASSO
Enumerator FE	No	No	Yes	No	Yes	Yes	YES
Taxi Terminal FE	No	No	No	Yes	No	No	No

Standard errors in parenthesis.

Columns (1) - (4) report VSL estimates obtained by estimating a regression of willingness to pay on the empirical risk reduction offered by a helmet. Column (5) estimates VSL as valuation normalized by the mortality risk reduction offered by a helmet. The mean VSL values and standard errors of the mean are reported in this column. In Panel A, the estimated helmet effectiveness from Liu et al. (2008) is used in constructing empirical estimates, while in Panel B the estimated effectiveness of helmets in Thailand from Ouellet and Kasantikul (2006) is used. Column (6) is the preferred VSL estimate and replicated for comparison. Column (7) re-weights the non-interacted experimental estimate to calculate the same LATE as the observational estimate reported in column (3). Experimental estimates in columns (6) - (7) report weak instrument robust confidence sets constructed using inversion of a conditional likelihood ratio (Moreira, 2003). All models exclude 35 observations for which surveyors reported motorcycle taxi drivers pretending to be passengers.

Table 9: Reduced form effect of information on demand

	(1)	(2)	(3)	(4)	(5)	(6)
	Individual treatments				Pooled treatments	
	WTP	WTP	log(WTP)	log(WTP)	WTP	WTP
Pure control	2.24	1.99	0.04	0.02		
	(4.58)	(4.49)	(0.10)	(0.10)		
Low treatment	-2.16	-2.23	-0.07	-0.08	-3.30	-1.99
	(1.90)	(1.84)	(0.05)	(0.05)	(1.59)	(1.79)
High treatment	1.65	1.95	-0.01	-0.01		
	(2.32)	(2.28)	(0.05)	(0.05)		
Low x BL Belief						-27.65
						(25.04)
Control mean	33.28	33.28			34.35	34.15
Pr(Low = High)	0.08	0.05	0.28	0.13		
Observations	1,536	1,536	1,536	1,536	1,536	1,427
Controls	None	LASSO	None	LASSO	LASSO	LASSO
Enumerator FE	No	Yes	No	Yes	Yes	Yes

Robust standard errors in parenthesis.

This table reports reduced form regressions of willingness to pay for a helmet, in PPP 2022 USD, on the information treatment that the respondent received. All models include wave fixed effects. The pure control was assigned with a probability of 0.1 in the first wave and 0 in the second wave of the experiment. The other treatment arms were assigned with equal probability. Columns 1-4 examine individual treatment assignments. Columns 5-6 pool control and high treatment observations, which had similar beliefs about the effectiveness of helmets. All models exclude 35 observations for which surveyors reported motorcycle taxi drivers pretending to be passengers.

Appendix

A Belief elicitation

The survey consisted of five modules. First, we collected tracking data and detailed demographic information about respondents. Second, we obtained information about their use of motorcycle taxis, including weekly ridership volume, standard trip length, trip types (e.g. commuting vs leisure), and reasons for using motorcycle taxis versus other modes of transportation (e.g. speed or cost). Third, surveyors elicited the respondents beliefs about their likelihood of dying in a motorcycle taxi accident. The first part of this module was identical across the treatment and control arms. We refer to variables measured during this portion of the module as baseline beliefs since the survey was identical across arms through this point (excluding the pure control), and the surveyors were not yet informed about treatment assignment. During the second component, surveyors presented empirical risk estimates to respondents in the treatments arms and elicited posteriors beliefs. Fourth, surveyors presented individuals in the treatment arms with the results of the appropriate study about helmet efficacy and then measured posterior beliefs about the effectiveness of helmets across the control and treatment arms. Finally, respondents completed the BDM game and received a helmet or mobile money payment. The third and fourth modules (those that asked questions about the risks of motorcycles and perceived safety benefits of helmets) were skipped for those in the pure control group.

Prior to the first survey wave, surveyors completed a week long pilot that was focused on identifying a reliable survey module to measure beliefs about mortality risk. The final set of questions begins by providing the passengers with reference points to help them express and contextualize rare events. We informed the respondents that Jamhuri (Independence) Day occurs one out of every 365.25 days and that a leap day occurs one out of every 1,461 days. These events were chosen because all individuals experience them with the same frequency, whereas indexing to something like the prevalence of HIV deaths would be correlated with socioeconomic status.

Surveyors next asked respondents how many fatal accidents occur per 100,000 motorcycle taxi trips. In addition to measuring the respondent's views about per trip risks, this question was designed to help the passengers think carefully about mortality risks by walking them through first thinking about how dangerous each trip is, then about their volume of trips. We then asked the respondent how many deaths there are per 10,000 passengers over the course of 1 year and 5 years in Nairobi. We asked over two time horizons to help the respondent think through risks incrementally and as a method of testing data quality since the 5 year response should be larger than the 1 year response.

We next asked the respondents about their own risk of suffering a fatal accident over the following 5 years. We did this in two steps. First, we asked them to select which range of risks seemed most accurate, for instance less than 1 in 10,000,000, between 1 in 10,000,000 and 1 in 1,000,000, etc. After selecting a range, we asked the passengers to respond with their exact belief within the range. Piloting revealed that this two step approach helped respondents answer accurately. A design feature of this survey is that the questions about one's own risk assessment and the number of deaths per 10,000 passengers over 5 years are asked in different ways but measure a similar outcome. The first question fixes the denominator and asks for a numerator, whereas the question about one's own risk asks for the risk in terms of a fraction, percent or decimal. These two questions need not perfectly align since the first considers an average Kenyan whereas the second asks about the respondent's own risk, but we may test whether they are broadly consistent to verify the quality of the measurements.

The respondents were then asked which information sources they used to construct their beliefs and whether they had been in a previous accident. We then presented those in the treatment groups with empirical estimates of their 5-year fatal accident risk as a function of their ridership and elicited posterior beliefs.

The motorcycle taxi context was chosen in part because empirical risks are high enough over the lifespan of a helmet to limit problems associated with understanding small probabilities. For instance, for a passenger that takes 6 trips per week on average, the median in this sample, we estimate that there is over a 1 in 5,000 chance that they will die in a motorcycle taxi accident in the next 5 years absent a helmet. Given limitations of the data used to construct this estimate, this may be a lower bound on the true risk. While this is still a relatively low probability, it is among the most probable causes of death for young adults and thus minimizes the cognitive burden of understanding small probabilities compared to other settings.

The efficacy of helmets are much easier to understand and communicate, so we follow a simpler survey procedure than that used to elicit beliefs about the risk of a fatal accident without a helmet. We first present the low treatment group with the Liu et al. (2008) estimate that helmets reduce one's likelihood of dying by 42% and the high treatment group with the Ouellet and Kasantikul (2006) estimate that helmets reduce mortality risk by 70%. Surveyors communicated this information by stating that for every 100 individuals that would die if no one wore a helmet, the study estimates that 42 (or 70) would survive if all had worn a helmet. The control and treatment groups were then asked for their own beliefs about how effectively helmets prevent death, expressed as the number of people that they believe would survive if all passengers wore a helmet per 100 deaths if no one wore a helmet.

B VSL Inference

My primary estimates of VSL and VSLY report homoskedastic standard errors. This analytic choice is supported by the latent utility model presented in section 3 since random assignment of T_i guarantees that it is independent of unobserved determinants of utility. Recall that the two-stage least squares model which identifies VSL is given by

$$v_i = \zeta_h + VSL\Delta r_i + X_i'\gamma_0 + \gamma_1 r_{0,i} + \epsilon_i$$

$$\Delta r_i = Z_i'\pi + X_i'\pi_c + \pi_r r_{0,i} + \nu_i$$

where ϵ_i denotes determinants of an agent's utility from a helmet which are observed to the individual but not the econometrician. In the case where $Z_i = T_i$, by randomization we know immediately that $\mathbb{E}[\epsilon_i^2|Z_i] = \mathbb{E}[\epsilon_i^2]$.

If $Z_i = (T_i', r_{0,i} \cdot T_i')'$, then errors may be heteroskedastic with respect to $r_{0,i}$. However, controls for $r_{0,i}$ will ensure homoskedasticity because $r_{0,i} \cdot T_i$ adds no information about ϵ_i^2 after accounting for $r_{0,i}$, so homoskedastic standard errors about VSL will still be accurate.

Formally, fix $r_{0,i}$. If $r_{0,i} \neq 0$, then T_i is uniquely determined by $r_{0,i}$ and $T_i \cdot r_{0,i}$ so $\mathbb{E}[\epsilon_i^2|r_{0,i} \cdot T_i, r_{0,i}] = \mathbb{E}[\epsilon_i^2|T_i, r_{0,i}] = \mathbb{E}[\epsilon_i^2|r_{0,i}]$ by the independence of T_i.

If $r_{0,i} = 0$, then T_i is not restricted by $r_{0,i} \cdot T_i$ so immediately$\mathbb{E}[\epsilon_i^2|r_{0,i} \cdot T_i, r_{0,i}] = \mathbb{E}[\epsilon_i^2|r_{0,i}]$. Denote $\sigma^2(r_{0,i}) = \mathbb{E}[\epsilon_i^2|r_{0,i}]$.

Keeping $r_{0,i}$ fixed, we have from the asymptotic variance formula for two-stage least squares that

$$\begin{aligned}
Avar(\sqrt{N}(\widehat{VSL} - VSL|r_{0,i})) &= \mathbb{E}[(\Delta\hat{r}_i)^2]^{-1} plim \frac{\Delta r Z(Z'Z)^{-1}Z'\epsilon\epsilon' Z(Z'Z)^{-1}Z'\Delta r}{N}\,\mathbb{E}[(\Delta\hat{r}_i)^2]^{-1} \\
&= \mathbb{E}[(\Delta\hat{r}_i)^2]^{-1}\sigma^2(r_{0,i})\,\mathbb{E}[(\Delta\hat{r}_i)^2]\,\mathbb{E}[(\Delta\hat{r}_i)^2]^{-1} \\
&= \epsilon_i^2(r_{0,i})\,\mathbb{E}[(\Delta\hat{r}_i)^2]^{-1} := V(r_{0,i})
\end{aligned}$$

where $\Delta\hat{r}_i = P_Z \Delta r_i$.

Hence, by the Law of Total Variance,

$$Avar(\sqrt{N}(\widehat{VSL} - VSL)) = E[V(r_{0,i})] + Var(\mathbb{E}[\sqrt{N}(\widehat{VSL} - VSL)|r_{0,i}])$$

$$= E[V(r_{0,i})] = \mathbb{E}[\epsilon_i]^2 \, \mathbb{E}[(\Delta\hat{r}_i)^2]^{-1}$$

Since errors are homoskedastic, two-stage least squares is efficient among linear IV estimators. I report weak instrument robust confidence sets alongside homoskedastic standard errors in the primary tables to demonstrate robustness against weak instruments. I use the Stata package *weakiv* (Finlay et al., 2016) to construct the confidence sets and use CLR inversion for over-identified models (Moreira, 2003) and AR inversion for just identified models (Anderson and Rubin, 1949). These confidence sets were selected for efficiency in the respective cases (Andrews et al., 2007).

Randomization was conducted at the individual level. As a result, clustering standard errors is not necessary for inference over the sampled population. However, I report two-stage cluster bootstrap (TSCB) standard errors and confidence intervals over the full sample in Appendix Table A4. I use 1,000 bootstrap iterations.

I report TSCB rather than analytic clustered standard errors because Abadie et al. (2022) demonstrate that clustered standard errors are too conservative in cases where randomization is assigned at the individual level, and the authors demonstrate that correlation between residuals within clusters does not necessitate clustering given this treatment assignment mechanism. Furthermore, the TSCB, which is based on study design, captures clustered sampling when the number of sampled clusters is a large share of the population number of clusters which accurately describes this setting. I report boostrapped standard errors because Abadie et al. (2022) do not derive analytic solutions for instrumental variable regression and bootstrapped confidence intervals are not symmetrical: there is more upward than downward uncertainty in VSL estimates.

C Appendix Tables

A1: Non-response rates and balance: Demographics

	(1) Control Summary statistics	(2) Pure control - control	(3) Low treatment - Control	(4) High treatment - Control	(5) High treatment - low treatment
Age	0.000	0.000	-0.000	0.004*	0.004
	[0.000]	(0.006)	(0.002)	(0.002)	(0.003)
Female	0.000	0.000	0.000	0.000	0.000
	[0.000]	(0.000)	(0.000)	(0.000)	(0.000)
Health (1-5)	0.000	0.000	0.000	0.000	0.000
	[0.000]	(0.000)	(0.000)	(0.000)	(0.000)
Life expectancy	0.015	-0.016	0.003	0.004	0.000
	[0.124]	(0.018)	(0.008)	(0.008)	(0.009)
Employed	0.000	0.000	0.000	0.000	0.000
	[0.000]	(0.000)	(0.000)	(0.000)	(0.000)
Wage (Ksh/hr)	0.111	0.055	0.042*	0.059**	0.017
	[0.314]	(0.045)	(0.023)	(0.023)	(0.023)
$\mathbb{E}$[Wage in 5 years/Wage today]	0.115	0.047	0.043*	0.058**	0.015
	[0.319]	(0.046)	(0.023)	(0.024)	(0.024)
1(children)	0.000	0.000	0.000	0.000	0.000
	[0.000]	(0.000)	(0.000)	(0.000)	(0.000)
Digit span recall	0.000	0.000	0.000	0.000	0.000
	[0.000]	(0.000)	(0.000)	(0.000)	(0.000)
Years of education	0.007	-0.008	0.001	-0.002	-0.003
	[0.081]	(0.010)	(0.005)	(0.005)	(0.005)
1(primary school complete)	0.007	-0.008	0.001	-0.002	-0.003
	[0.081]	(0.010)	(0.005)	(0.005)	(0.005)
1(secondary school complete)	0.007	-0.008	0.001	-0.002	-0.003
	[0.081]	(0.010)	(0.005)	(0.005)	(0.005)
1(college degree)	0.007	-0.008	0.001	-0.002	-0.003
	[0.081]	(0.010)	(0.005)	(0.005)	(0.005)
N (exc. control in col. 2-4)	452	80	531	473	
p-value of joint orthogonality		0.271	0.416	0.138	0.652

Standard deviations in brackets. Standard errors in parenthesis.

Column (1) reports the mean and standard deviation of the indicated variable across the full sample. Columns (2) - (5) report the difference in mean of each variable across treatment arms, where terms in parentheses are standard errors of this pairwise t-test. Missing values are imputed using the median across the full sample.

A2: Non-response rates and balance: Motorcycle taxi use

	(1) Control Summary statistics	(2) Pure control - control	(3) Low treatment - Control	(4) High treatment - Control	(5) High treatment - low treatment
Trips/week	0.000	0.000	0.000	0.000	0.000
	[0.000]	(0.000)	(0.000)	(0.000)	(0.000)
Average trip length (minutes)	0.000	0.000	0.000	0.000	0.000
	[0.000]	(0.000)	(0.000)	(0.000)	(0.000)
$\mathbb{E}$[Deaths/10,000 passengers over 5 years]	0.004	NA	0.001	0.006	0.005
	[0.066]	NA	(0.005)	(0.005)	(0.006)
Baseline belief: 10000*Pr(Fatal accident over 5 years)	0.018	NA	0.001	0.006	0.004
	[0.132]	NA	(0.009)	(0.009)	(0.009)
Motorcycle trip types	0.000	0.000	0.000	0.000	0.000
	[0.000]	(0.000)	(0.000)	(0.000)	(0.000)
Reasons for using motorcycles	0.000	0.000	0.000	0.000	0.000
	[0.000]	(0.000)	(0.000)	(0.000)	(0.000)
Risk information sources	0.000	NA	0.000	0.000	0.000
	[0.000]	NA	(0.000)	(0.000)	(0.000)
N (exc. control in col. 2-4)	452	80	531	473	
p-value of joint orthogonality		.	0.962	0.554	0.652

Standard deviations in brackets. Standard errors in parenthesis.

Column (1) reports the mean and standard deviation of the indicated variable across the full sample. Columns (2) - (5) report the difference in mean of each variable across treatment arms, where terms in parentheses are standard errors of this pairwise t-test. Missing values are imputed using the median across the full sample.

A3: Effect of information on beliefs: Estimated including manipulated surveys

	(1) Posterior risk	(2) Posterior risk	(3) Helmet effectiveness	(4) Helmet effectiveness	(5) Risk reduction	(6) Risk reduction, winsorized
Low treatment	-20.56	-20.83	-14.06	-14.25	-77.09	-41.52
	(21.72)	(21.93)	(0.97)	(0.99)	(18.46)	(14.28)
High treatment	7.58	5.88	-4.12	-4.54	-28.33	0.66
	(24.44)	(24.86)	(0.87)	(0.92)	(19.35)	(15.87)
Control mean	339.71	339.71	78.85	78.85	226.92	233.96
Pr(High treatment = low treatment)	0.13	0.15	0.00	0.00	0.00	0.00
Observations	1,455	1,455	1,457	1,457	1,455	1,455
Controls	BL Risk	LASSO	BL Risk	LASSO	LASSO	LASSO
Enumerator FE	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors in parenthesis.

Columns (1) - (2) report 10,000 times the respondent's stated posterior belief about the probability that they will suffer a fatal accident over the next 5 years, the lifespan of a helmet, if they do not purchase a helmet. Columns (3) - (4) report the respondent's stated belief about the effectiveness of motorcycle helmets, expressed as the percentage reduction in mortality risk. Columns (5) - (6) report 10,000 times the likelihood that a helmet will save the respondent's life over its 5 year lifespan based on their beliefs. The outcome in column (6) is winsorized at the 2nd and 98th percentiles. All estimates include wave fixed effects. Respondents assigned to the "Low treatment" group received the results of Liu et al. (2008) which estimates that helmets are 42% effective. Those assigned to "High treatment" were presented with the results of Ouellet and Kasantikul (2006) that helmets were 70% effective in a study in Thailand. Individuals in both treatment arms received empirical estimates of their risk of a fatal accident based on ridership. These estimates include 35 observations for which surveyors reported motorcycle taxi drivers pretending to be passengers.

A4: Value of a statistical life: Estimates with two-stage cluster bootstrap standard errors

	Raw WTP				Winsorized WTP	
	(1) Interacted	(2) Treatment only	(3) Interacted	(4) Treatment only	(5) Interacted	(6) Treatment only
VSL	210.95	340.36	219.45	301.76	97.15	245.40
	(141.19)	(398.99)	(398.99)	(184.07)	(469.28)	(71.25)
Cragg-Donald F-stat	40.12	10.99	40.75	10.63	40.75	10.63
95% CI	[92.89, 418.35]	[136.45, 1,065.20]	[136.45, 1,065.20]	[-136.29, 282.16]	[-288.97, 729.01]	[70.62, 224.63]
Observations	1,425	1,425	1,425	1,425	1,425	1,425
Controls	BL Risk	BL Risk	LASSO	LASSO	LASSO	LASSO
Enumerator FE	Yes	Yes	Yes	Yes	Yes	Yes

Two-stage cluster bootstrap standard errors in parenthesis.

This table reports VSL estimates with inference conducted using the two-stage cluster bootstrap (TSCB) from Abadie et al. (2022) constructed using 1,000 iterations. TSCB confidence intervals are included in brackets. Columns labeled “Interacted” use a first stage consisting of treatment indicators and treatment indicators interacted with baseline motorcycle risk beliefs as instruments for the mortality risk reduction of a helmet. Columns labeled “Treatment only” use only treatment assignment as an instrument. All models control for baseline risk beliefs. Columns (5) - (6) report estimates where willingness to pay is winsorized at the 2nd and 98th percentiles. These estimates exclude 35 observations for which surveyors reported motorcycle taxi drivers pretending to be passengers.

A5: Value of a statistical life: Estimates including manipulated surveys

Panel A: Full sample

	Raw WTP				Winsorized WTP	
	(1) Interacted	(2) Treatment only	(3) Interacted	(4) Treatment only	(5) Interacted	(6) Treatment only
VSL	39.07	268.22	57.90	206.96	43.64	227.79
	(96.89)	(273.05)	(97.09)	(277.23)	(69.06)	(200.64)
Cragg-Donald F-stat	41.88	9.99	42.05	9.57	42.05	9.57
Weak IV Robust Confidence Set	[-156.16, 235.71]	[-291.49, 1,033.81]	[-156.16, 235.71]	[-387.95, 964.79]	[-94.28, 183.40]	[-174.68, 755.33]
Inversion test	CLR	CLR	CLR	CLR	CLR	CLR
Observations	1,455	1,455	1,455	1,455	1,455	1,455
Controls	BL Risk	BL Risk	LASSO	LASSO	LASSO	LASSO
Enumerator FE	Yes	Yes	Yes	Yes	Yes	Yes

Panel B: Treated respondents only

	Raw WTP				Winsorized WTP	
	(1) Interacted	(2) Treatment only	(3) Interacted	(4) Treatment only	(5) Interacted	(6) Treatment only
VSL	28.70	746.72	-21.01	673.50	28.45	364.72
	(246.25)	(523.41)	(251.32)	(539.06)	(171.66)	(352.54)
Cragg-Donald F-stat	17.18	9.63	15.94	8.46	15.94	8.46
Weak IV Robust Confidence Set	[-506.22, 597.35]	[-155.47, 2,673.76]	[-578.80, 544.73]	[-284.18, 2,752.97]	[-338.36, 409.58]	[-309.98, 1,627.17]
Inversion test	CLR	AR	CLR	AR	CLR	AR
Observations	1,006	1,006	1,006	1,006	1,006	1,006
Controls	BL Risk	BL Risk	LASSO	LASSO	LASSO	LASSO
Enumerator FE	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parenthesis.

Panel A reports regressions using observations in the control group, low treatment group, and high treatment group. Panel B uses observations from the low and high treatment arms only. Columns labeled "Interacted" use a first stage consisting of treatment indicators and treatment indicators interacted with baseline motorcycle risk beliefs as instruments for the mortality risk reduction of a helmet. Columns labeled "Treatment only" use only treatment assignment as an instrument. All models control for baseline risk beliefs. Columns (5) - (6) report estimates where willingness to pay is winsorized at the 2nd and 98th percentiles. Over-identified models report weak instrument robust confidence sets constructed using inversion of a conditional likelihood ratio (Moreira, 2003). For just-identified models, weak instrument robust confidence sets are constructed by inverting an AR test (Anderson and Rubin, 1949; Chernozhukov and Hansen, 2008). These estimates include 35 observations for which surveyors reported motorcycle taxi drivers pretending to be passengers.

A6: Robustness of VSL to alternative assumptions

Panel A: Change in planned future ridership

	Raw WTP				Winsorized WTP	
	(1) Interacted	(2) Treatment only	(3) Interacted	(4) Treatment only	(5) Interacted	(6) Treatment only
VSL	138.75	281.75	145.10	249.62	61.25	189.36
	(63.62)	(183.57)	(64.49)	(190.75)	(47.12)	(141.19)
Cragg-Donald F-stat	45.37	10.37	44.85	9.58	44.85	9.57
Weak IV Robust Confidence Set	[13.39, 269.07]	[-80.45, 787.66]	[18.40, 277.56]	[-138.62, 781.42]	[-32.24, 156.65]	[-91.29, 557.53]
Inversion test	CLR	CLR	CLR	CLR	CLR	CLR
Observations	1,425	1,425	1,425	1,425	1,425	1,425
Controls	BL Risk	BL Risk	LASSO	LASSO	LASSO	LASSO
Enumerator FE	Yes	Yes	Yes	Yes	Yes	Yes

Panel B: Different beliefs about helmet lifespan

	Raw WTP				Winsorized WTP	
	(1) Interacted	(2) Treatment only	(3) Interacted	(4) Treatment only	(5) Interacted	(6) Treatment only
VSL	250.69	325.53	266.17	283.10	137.50	229.37
	(87.84)	(240.24)	(87.45)	(241.69)	(63.32)	(178.38)
Cragg-Donald F-stat	41.49	10.38	42.07	10.05	42.06	10.04
Weak IV Robust Confidence Set	[82.06, 436.41]	[-148.15, 1,011.89]	[99.14, 451.69]	[-207.25, 958.28]	[14.09, 268.43]	[-124.80, 690.24]
Inversion test	CLR	CLR	CLR	CLR	CLR	CLR
Observations	1,425	1,425	1,425	1,425	1,425	1,425
Controls	BL Risk	BL Risk	LASSO	LASSO	LASSO	LASSO
Enumerator FE	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parenthesis.

Panel A considers VSL estimates that account for increases in planned future ridership associated with receiving a helmet. For those that received a cash gift, this value is imputed by regressing planned future ridership on past ridership fully interacted with treatment assignment. Panel B considers VSL estimates using the risk reduction offered by a helmet over the respondent's stated belief about the lifespan of the helmet, rather than the manufacturers suggestion. All columns use the full sample and a first stage consisting of treatment indicators and treatment indicators interacted with baseline motorcycle risk beliefs as instruments for the mortality risk reduction of a helmet. All models control for baseline beliefs. I report weak instrument robust confidence sets constructed using inversion of a conditional likelihood ratio (Moreira, 2003). All models exclude 35 observations for which surveyors reported motorcycle taxi drivers pretending to be passengers.

A7: Value of a statistical life: Winsorized beliefs

Panel A: Full sample

	Raw WTP				Winsorized WTP	
	(1) Interacted	(2) Treatment only	(3) Interacted	(4) Treatment only	(5) Interacted	(6) Treatment only
VSL	524.25	658.62	589.91	643.49	276.59	428.28
	(247.78)	(408.41)	(276.14)	(443.88)	(193.24)	(319.68)
Cragg-Donald F-stat	9.54	7.38	7.94	6.23	7.94	6.23
Weak IV Robust Confidence Set	[62.94, 1,173.32]	[-89.01, 1,986.66]	[62.94, 1,173.32]	[-180.78, 2,259.42]	[-109.06, 762.20]	[-190.80, 1,463.46]
Inversion test	CLR	CLR	CLR	CLR	CLR	CLR
Observations	1,425	1,425	1,425	1,425	1,425	1,425
Controls	BL Risk	BL Risk	LASSO	LASSO	LASSO	LASSO
Enumerator FE	Yes	Yes	Yes	Yes	Yes	Yes

Panel B: Treated respondents only

	Raw WTP				Winsorized WTP	
	(1) Interacted	(2) Treatment only	(3) Interacted	(4) Treatment only	(5) Interacted	(6) Treatment only
VSL	663.15	950.18	672.10	988.96	313.34	515.44
	(393.71)	(583.39)	(404.58)	(615.16)	(275.29)	(408.82)
Cragg-Donald F-stat	10.67	10.97	10.00	9.93	10.00	9.93
Weak IV Robust Confidence Set	[-64.74, 1,749.68]	[-48.50, 2,953.06]	[-74.07, 1,816.70]	[-52.24, 3,240.08]	[-235.11, 1,036.33]	[-240.50, 1,884.87]
Inversion test	CLR	AR	CLR	AR	CLR	AR
Observations	982	982	982	982	982	982
Controls	BL Risk	BL Risk	LASSO	LASSO	LASSO	LASSO
Enumerator FE	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parenthesis.

Panel A reports regressions using observations in the control group, low treatment group, and high treatment group. Panel B uses observations from the low and high treatment arms only. Columns labeled “Interacted” use a first stage consisting of treatment indicators and treatment indicators interacted with baseline motorcycle risk beliefs as instruments for the mortality risk reduction of a helmet. Columns labeled “Treatment only” use only treatment assignment as an instrument. All models control for baseline risk beliefs. Columns (5) - (6) report estimates where willingness to pay is winsorized at the 2nd and 98th percentiles. Over-identified models report weak instrument robust confidence sets constructed using inversion of a conditional likelihood ratio (Moreira, 2003). For just-identified models, weak instrument robust confidence sets are constructed by inverting an AR test (Anderson and Rubin, 1949; Chernozhukov and Hansen, 2008). All models exclude 35 observations for which surveyors reported motorcycle taxi drivers pretending to be passengers. Beliefs about the mortality reduction from a helmet are winsorized at the 2nd and 98th percentiles.