A New Experimental Method for Estimating Demand for Non-market Goods With an Application to the Value of a Statistical Life*

Grady Killeen

University of California, Berkeley

March 10, 2025

Abstract

Economists often study consumer preferences for non-market goods such as clean air and health. This paper introduces a new revealed preference method to estimate demand for these amenities and applies it to study preferences for mortality risk reduction, the value of a statistical life (VSL), in Kenya. My approach is to update beliefs about the efficacy of a product that reduces mortality risk (a new helmet), and elicit product choice. This generates instruments allowing one to use subjective belief data to estimate demand, avoiding the need to assume rational expectations. This procedure does not require beliefs to be reported without error, it only makes the assumption that measurement error is classical. My design validates this assumption using multiple instruments and insights from local average treatment effects. This method yields estimates of VSL of $224, near the left tail of East African estimates. I find that existing methods for estimating VSL produce skewed estimates of the parameter, driven by severe violations of rational expectations. These findings help rationalize low demand for health products in low and middle-income countries. The VSL estimates also change the conclusion of 4 out of 5 recent benefit-cost analyses and imply that billions of dollars of development aid may be misallocated, highlighting the importance of accurate estimates of demand for non-market amenities. (*JEL* O18, R49, J17)

*I thank Edward Miguel and Supreet Kaur for excellent feedback and advising. This project benefited from suggestions by Kurt Lavetti, Jeremy Magruder, Kelsey Jack, Marco Gonzalez-Navarro, Sam Wang, Elif Tasar, Catherine Che, Matteo Saccararola, Nick Shankar, Nicholas Swanson, Michael Walker and attendants of the UC Berkeley Development Lunch and the University of Chicago Causal Inference Conference. I thank William Jack, Whitney Tate, Nyambaga Muyesu, Josephine Okello and the Georgetown University Initiative on Innovation, Development, and Evaluation for support with implementation. 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

Many goods that are important to welfare – such as clean air and water, health, privacy and reputation – are not directly traded. Pricing these amenities is important to design public policies, but the absence of markets makes it difficult to identify consumer preferences. This paper introduces a new experimental method to estimate demand for non-market goods by revealed preference and applies it to estimate consumers' willingness to pay to reduce mortality risk. My approach draws on the fact that (even though these amenities are not traded), there are often consumer products which change exposure to or consumption of such goods (e.g. air or water purifiers).

Attempting to exogenously vary exposure to non-market goods is often infeasible. Rather, my approach shifts beliefs about the extent to which a given consumer product affects exposure to an amenity. Demand for the consumer product then enables an inference about how the amenity affects consumer utility. I estimate demand for the non-market good by examining how willingness to pay for the product changes with perceived exposure to the amenity, using the randomized provision of information to instrument for subjective beliefs. This method does not assume rational expectations, a common requirement of traditional approaches. I also do not require that agents fully update beliefs in response to information, that their beliefs adhere to a specific learning model, or that individuals report expectations without error. The primary identifying assumption is that measurement error in reported beliefs is classical.

I design two tests to validate this assumption. One leverages multiple information treatments of varying intensity to verify that estimates scale with changes in beliefs, helping rule out features like experimenter demand effects. The second adapts insights from the local average treatment effects literature to construct estimates that weight observations differently as a function of priors. This identifies bias driven by common types of belief misreporting such as rounding or "S-shaped" probability weighting (Kahneman and Tversky, 1979). These tests are useful in their own right since they generalize to other settings where researchers rely on data from elicited beliefs.

I apply this method to study demand for mortality risk reduction, the value of a statistical life (VSL), in Kenya. VSL is important for public policy but notoriously challenging to estimate. Mortality risk is a feature of economic decisions including job choice, healthcare, environmental regulation and transportation. Policymakers at agencies such as the Environmental Protection Agency directly use VSL in benefit-cost analyses, and NGOs rely on the parameter to allocate billions of dollars of development aid.[1] In theory, VSL estimates allow practitioners to use consumers' own preferences for trading off mortality risk and consumption to guide policy, maximizing welfare and avoiding paternalistic judgments. But economists have raised concerns that methods typically used to estimate the parameter may be prone to bias, causing misallocation (Ashenfelter, 2006).

VSL is often estimated by taking decisions where mortality risk is salient, then estimating demand (or supply) models where the actual level of mortality risk present in the data is used as a proxy for beliefs.[2] This approach is biased if agents have beliefs about mortality risk that differ from the econometrician's estimate (i.e. they do not have rational expectations). Moreover, estimates may be prone to omitted variable bias if agents select into dangerous behaviors (Ashenfelter, 2006; Ashenfelter and Greenstone, 2004). VSL may also be heterogeneous, and many studies estimate it over selected populations where there is quasi-random variation in mortality risk (Greenberg et al., 2021). Consistent with potential bias, estimates of VSL vary by orders of magnitude: in Africa, they range from about \$0-\$700,000 (Berry et al., 2020; León and Miguel, 2017).[3]

In this paper, I estimate the VSL of urban Kenyans by presenting motorcycle taxi passengers with randomly assigned information about the efficacy of helmets at preventing death. I then offer them the choice between a helmet or cash. As previously discussed, this approach does not rely on rational expectations. Since estimates are identified from the random assignment of information, selection into risk is also not a concern. Motorcycle taxi passengers were assigned to a control group (receiving no information) or one of two treatments. One treatment arm was presented with personalized estimates of unhelmeted mortality risk and the results of Liu et al. (2008), which estimates that helmets reduce one's risk of dying by 42%. The second received the same information about motorcycle risks, but the results of Ouellet and Kasantikul (2006), which estimates that helmets are 70% effective. I then elicited respondents' posteriors about the likelihood that a helmet would save their life and estimated helmet demand using a Becker et al. (BDM, 1964) mechanism modified to reduce liquidity constraints. I estimate VSL by fitting an

[1] For instance, GiveWell uses "moral weights" when ranking charities, which are derived from VSL as discussed in section 5.1. GiveWell directed more than \$500 million to charities in 2021 alone.

[2] León and Miguel (2017) estimate demand for different modes of transportation with high mortality risk and Greenberg et al. (2021) consider reenlistment decisions of soldiers.

[3] An approach that does not assume rational expectations is stated preference VSL elicitation which directly asks agents how much they would pay for hypothetical mortality reductions. Agents have no incentivize to answer truthfully and these estimates may be sensitive to social desirability bias, so I focus on revealed preference methods.

instrumental variables regression of willingness to pay on the probability that a helmet will save the agent's life, using treatment assignment to instrument for beliefs.

I study the Kenyan motorcycle taxi market because most consumers have yet to adopt this product, which would substantially reduce their mortality risk, allowing one to estimate VSL over a minimally selected population. Traffic accidents are a leading cause of death in East Africa, so helmets are one of the most important safety investments that motorcycle users can make. Motorcycle taxi use is widespread in Kenya, but only an estimated 3% of passengers wore helmets near the time of the study (Bachani et al., 2017).[4] Since a passenger's empirical likelihood of a fatal accident is a function of trip volume, this setting also allows me to compare my approach to VSL estimates from traditional methods.

In the first main finding of the study, I estimate that urban Kenyans have low average demand for safety, with a VSL of only PPP USD $224 (about 5% of median annual household income).[5] Exposure to the results of Liu et al. (2008) reduces average beliefs about the likelihood that a helmet will save an agent's life (a product of beliefs about the mortality risk of motorcycles and helmet efficacy) from about 2.22% to 1.41%. This reduces willingness to pay for a helmet by about $3. Despite low demand for safety, I recover informative bounds on VSL, rejecting values below $34 and above $429 with 95% confidence.

My method also produces estimates of VSL heterogeneity consistent with theory. I estimate that the parameter is about $500 higher among those with above median wealth, an elasticity with respect to income of 1.14, and I find that healthier agents have higher VSLs. Theory predicts that VSL should rise rapidly with income because it is inversely proportional to the marginal utility of consumption. This insight can be used to estimate how rapidly the value of marginal consumption would need to fall with income to rationalize different VSL estimates across income levels. I apply this exercise to León and Miguel (2017), which estimate a VSL over $700,000 (adjusting for inflation) on a population with an average income over $75,000 (compared to $5,000 in this sample). This implies utility curvature that aligns closely with published values, meaning that theory predicts the order of magnitude differences between my VSL estimates and those from

[4] Tatah et al. (2023) document that a minority of commuting trips occur via motorcycle, but most individuals also use motorcycle taxis for last mile journeys or other trip types.

[5] 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.

high-income populations. These findings help rationalize the fact that LMIC consumers often have a low willingness to pay for health products such as malaria nets (Dupas and Miguel, 2017).

The second component of this study validates the identifying assumption that beliefs are measured without non-classical error. Two leading reasons this may fail are experimenter demand effects and systematically misreported beliefs. The study design incorporates two features to test for these (and other) confounds, and I present additional results which suggest that estimates are not likely biased. First, the study randomized whether agents receive no information, the results of Ouellet and Kasantikul (2006), or the results of Liu et al. (2008) so there are multiple signals that vary in intensity. If beliefs measures have only classical error, then VSL estimates should be unchanged if one of the treatment arms is excluded, but otherwise one would expect values to differ since posteriors are centered at different points across treatments. I show that any combination of arms yields similar estimates. Furthermore, incentivized willingness to pay for a helmet falls when agents are exposed to Liu et al. (2008), which is far from priors, indicating that it is unlikely that the results reflect demand effects where agents report a change in beliefs when none occurred.

The other design feature that allows for tests of biased beliefs is based on logic from the literature on local average treatment effects. Instrumenting for posteriors using treatment assignment weights observations linearly in priors, but one may construct an instrument that weights observations proportionally to priors squared, yielding similar results.[6] The similarity of estimates suggests that estimates are not likely affected by bias in posteriors that would be correlated with priors, which I show includes most cases where agents systematically under or over-report probabilities (e.g. "S-shaped" weighting or rounding low probabilities) and experimenter demand effects.

In additional checks, I demonstrate that helmet demand is similar between the control and a pure control group that was not asked questions about motorcycle risks or helmet benefits, that there is no detectable heterogeneity with respect to education or performance on a digit span recall test, and that distinct measures of beliefs are highly correlated. The study also presented treated respondents with information about the risks of motorcycles and efficacy of helmets, but agents only reported updating beliefs about helmets. VSL estimates are also similar if I instrument for beliefs using an observational belief shifter: an indicator for knowing a motorcycle accident victim.

[6]Point estimates are withing $100 of each other, and a J-test of over-identifying restrictions fails to reject the equality of the values.

The BDM game was implemented over a high-stakes and realistic decision since evidence indicates that the mechanism is reliable in similar settings (Cole et al., 2020; Berry et al., 2020).

The third part of the paper tests whether rational expectations, a core assumption required by most methods for estimating VSL, holds in this context. The data support the view that beliefs differ systematically from empirical probabilities because agents learn from their own experiences and those of their social network, not from representative data sources. Those that know someone involved in a motorcycle accident perceive their probability of suffering a fatal accident to be high, but those that ride motorcycles frequently do not perceive a higher 5-year mortality risk. Respondents also report twice the risk if it rained the day of the survey, raising the salience of risk. Deviations from rational expectations only add noise using the new method, but they are likely to bias estimates from typical approaches that proxy for beliefs with empirical risk. Consistent with this prediction, two common methods for estimating VSL produce results outside of confidence sets obtained from the new approach, and estimates differ significantly from each other.

These results have important policy implications. The VSL estimates are similar to the lower range of existing estimates from low and middle-income (LMIC) economies (Kremer et al., 2011; Berry et al., 2020), but they are much smaller than the values used in 5 recent benefit-cost analyses. In 4 of 5 cases, benefit-cost ratios fall below 1 if re-calculated with my estimates. In addition, an NGO that directs USD billions of aid allocates funds based on a VSL over 100 times as high. This suggests that practitioners may underweight recipient consumption gains, producing misallocation.

I caution, however, that the results do not imply that directing health resources to rich countries maximizes welfare. As discussed in the model, theory predicts that VSL will rise with income largely due to the falling opportunity cost of funds. It would therefore be incorrect to apply this value for targeting health resources with a fixed funding source (e.g. the WHO) since no LMIC consumers face a health-consumption trade-off. For instance, recent studies of vaccine production and allocation have discounted the value of lives saved in LMIC countries even though costs are borne predominantly by rich consumers (e.g. Ahuja et al., 2021; Agrawal et al., 2023). Theory and results suggest that this approach substantially under-estimates LMIC welfare gains. But it is accurate to use these values when deciding between distributing cash transfers or health assistance to consumers, or for policymaking by LMIC governments.

This paper advances the literature on VSL estimation by introducing and validating a method for estimating the parameter that does not rely on rational expectations and is robust to agents selecting into dangerous behaviors. To my knowledge, this is among the first studies to produce a precise estimate of VSL without assuming rational expectations and to test whether agents' beliefs satisfy the assumption.[7] This builds on work including Ashenfelter and Greenstone (2004) and Greenberg et al. (2021) which study how endogeneity and heterogeneity affect VSL measurements. It also relates to Baylis et al. (2023) which finds that deviations from rational expectations likely biases estimates of demand for air quality. The VSL estimates are also independently important because knowledge of demand for safety in LMICs is limited (Greenstone and Jack, 2015).[8] I produce one of the first urban estimates. In addition, the design ensures that results are not biased by cash on hand, a concern with incentivized methods in LMICs (Redfern et al., 2019).

The methodology introduced in this study could be valuable to amenities other than safety. Demand for non-market goods including time, privacy, reputation, and environmental quality are economically important but challenging to estimate, particularly in LMICs (Campbell and Brown, 2003; Greenstone and Jack, 2015). Policy decisions often rely on estimates from methods that are not incentivized (e.g. contingent valuation surveys) when the variation needed for revealed preference estimates is unavailable (Mendelsohn and Olmstead, 2009). Revealed preference methods also tend to assume rational expectations, which contradicts recent evidence (Baylis et al., 2023). The framework I introduce could be used to construct revealed preference estimates of demand for these amenities – without assuming rational expectations – by applying it to products such as ride sharing services, cyber security software, or air purifiers.

This research also contributes to a literature that elicits agents' subjective beliefs for use in estimating economic models (e.g. Manski, 2004; Wiswall and Zafar, 2018; Blass et al., 2010). This paper demonstrates that researchers may leverage information treatments to produce instruments for beliefs that overcome measurement error, which has prevented economists from using belief data in certain settings (León and Miguel, 2017). Furthermore, I show that one may test for leading sources of bias using multiple signals or multiple instruments. More practically, the survey design used to elicit beliefs in this setting was effective at minimizing rounding which could broadly improve measurement.

[7]Shrestha (2020) estimates VSL using different experimental methods, but confidence intervals are not informative.

[8]Kremer et al. (2011) and Berry et al. (2020) study the VSL of rural children, and León and Miguel (2017) and Ito and Zhang (2020) examine wealthy populations.

2 Study design and context

2.1 Motorcycle taxis and helmet use in Kenya

This study considers a sample of motorcycle taxi passengers in Nairobi, Kenya. This setting has three features which are ideal to study demand for mortality risk reduction. First, motorcycle taxis are near universally used and it is rare for taxi passengers to wear helmets, so VSL may be estimated over an informative sample. Second, motorcycle helmets substantially reduce one's risk of death, allowing one to estimate demand for a substantive improvement in safety. Third, empirical mortality risk may be estimated at the individual level since it is a function of ridership volume. This facilitates comparisons between different approaches for estimating VSL.

The use of motorcycle taxis is widespread and growing in East Africa, attributed to low costs and road congestion. In Kenya, there are an estimated 2.4 million drivers providing taxi services, combining for about 22 million trips per day.[9] Transportation is notoriously dangerous. Data from the National Transport and Safety Authority (NTSA) reports that 1,722 motorcycle drivers and passengers died in 2021, up from 715 in 2017.[10] Traffic accidents are the leading cause of death among boys 15-19 in Kenya, and a top-five cause of death for Kenyans aged 5-70.[11]

Despite the high risks of motorcycles, helmet use among taxi passengers is rare. Bachani et al. (2017) measured passenger helmet use at 3% in an observational study. At that time, helmeted passengers typically borrowed them from the driver. Anecdotal accounts suggest that consumers became concerned with the cleanliness of borrowed helmets during the COVID-19 pandemic. Consistent with this view, under 1% of motorcycle taxi passengers we approached reported regular helmet access. The low use of helmets suggests that demand may be low. However, the availability of effective helmets affordable to Kenyans is recent. The *FIA Foundation Safe and Affordable Helmet Program* began offering helmets in Kenya in 2021, about a year before the experiment was implemented, and a local producer began manufacturing helmets at about the same time. Retail diffusion of both products was limited outside of the Central Business District, and many consumers reported that they were unaware they could purchase a helmet separately from a motorcycle.

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

[10] Some sources estimate that official statistics underestimate deaths by a factor of 4. See for example *Nation.Africa*, "WHO: Kenya road deaths four times higher than NTSA reported," 2018.

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

One concern with this sample is that consumers with low demand for safety may select into riding motorcycles. This setting was chosen because most Kenyans use motorcycle taxis so selection is limited at the extensive margin. Motorcycle taxis are used similarly to traditional taxis in high-income countries. It is common to take one on trips where public transit is unavailable, for leisure or when running late, but less common to use them for daily commutes. Transportation surveys tend to focus on commuting, and I am not aware of representative data capturing the share of consumers that use motorcycle taxis. The Kenyan field team estimates that at least 85% of urban Kenyan adults use motorcycle taxis. Back of the envelope calculations also suggest that ridership needs to be high among urban adults to rationalize aggregate figures.[12] These figures are consistent with reports documenting that motorcycle taxis are a mainstream and widespread component of transportation (Nyachieo et al., 2024). Selection may be limited because alternative modes of public transportation are dangerous, crowded and uncomfortable (Habyarimana and Jack, 2011).

Unlike the extensive margin, ridership volume may be selected since many consumers do not use motorcycles for daily commute. The sampling was designed so that one's probability of being sampled is a known function of ridership, allowing for estimates that use re-weighting to examine how selection affects results. Section 5.1 presents details and shows that re-weighting is inconsequential, suggesting that selection on safety preferences is not a first order concern. The VSL of motorcycle taxi users is also policy relevant even if the population is not representative.

2.2 Recruitment

This study recruited consumers from motorcycle taxi stands in Nairobi during two waves of data collection. Surveyors censused 188 taxi stands and conducted surveys at 97, reaching 13 of Nairobi's 17 geographic constituencies. The stands were selected for broad geographic and demographic coverage. However, areas with high crime rates were excluded for surveyor safety. Survey locations are plotted over a map of Nairobi in Appendix Figure A1.

The study leveraged arrival times of consumers to sample from the population of passengers at each location. Surveyors attempted to recruit the first individual to arrive at a stand after completing a survey. Consumers that did not report regular access to a motorcycle helmet were informed that

[12]There were an estimated 22 million trips per day in 2022, and a population of 53 million. 40% of Kenyans are farmers and at least 10% are children 5 or younger. Respondents in our sample report taking about 7.5 trips per week on average. This would suggest that about 80% of urban Kenyans adults ride motorcycle taxis ($22 \approx 53 \cdot 0.5 \cdot \frac{7.5}{7} \cdot 0.78$).

they could choose a free helmet or a cash gift if they completed a 15-30 minute survey. The high value of the gifts (about \$15 on average) relative to survey time yielded a high response rate. Over 90% of passengers agreed to take part in the survey. The majority of those that did not participate lacked time, and under 1% were excluded because they regularly used a helmet.

Demographic information presented in Table 1 shows that the study reached a broad sample. This suggests that the VSL estimates are informative when considering other East African samples. Income aligns closely with representative samples of the population. 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 in 2021, and the Kenyan National Bureau of Statistics reported gross per capita production of \$7,907 for Nairobi county in 2017 (Nairobi is the richest county in Kenya).[13] However, the sample is not perfectly representative. More males (981) than females (444) were surveyed, explained by the fact that that men spend more time away from home. Education is relatively high, with an average of about 12 years of schooling completed.

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. There are four study arms: a pure control, control, and two treatments. The treatment groups were presented with information about the mortality risk of motorcycle taxis and the effectiveness of helmets at preventing death. The control arms received no information.

The pure control and control vary in the questions that they were asked. The pure control was not asked about motorcycle safety. In contrast, the control group was asked questions about the risks of motorcycles and their beliefs about the effectiveness of helmets. The pure control was included to test if asking respondents these questions affects demand. Pure control observations are excluded from VSL estimates since beliefs were not elicited. In practice, there are no differences in demand between the pure control and control groups.

The treatment arms were presented different studies estimating helmet effectiveness. Respondents first received an estimate of their unhelmeted mortality risk over the 5 year lifespan of a helmet.[14] They were then presented with one of two studies. Those in a "low treatment" group

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

[14] I calculated per trip mortality risk estimates for the average Kenyan from NTSA data, then estimated 5 year risks based on the respondent's ridership volume.

were presented 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), which estimates that high quality helmets reduce fatality risk by about 70% in Thailand. Surveyors followed standardized scripts to present the information, so only the study and results vary across arms. The sources of information were disclosed so respondents could judge the information's credibility.

Both studies are high quality so no respondents were given misleading information. There is a strong consensus that motorcycle helmets are effective, but there is uncertainty about exactly how well they work. In fact, 70% is approximately the upper bound of the 95% confidence interval reported in Liu et al. (2008).

2.4 Helmet valuations

The study measured demand for a helmet at the end of the survey using a Becker et al. (1964) willingness to accept mechanism. Respondents were asked the smallest cash payment that they would prefer to a free helmet. Surveyors then revealed a randomly selected payment amount between 5 and 600 Kenyan shillings (Ksh). If the payment amount was greater than or equal to the respondent's bid, then they received the cash. Otherwise they were given the free helmet. The study leveraged willingness to accept, rather than pay, to ensure that a lack of cash on hand did not prevent agents from acquiring helmets, which would bias VSL.

The maximum draw was Ksh 600 based on assessments by an NGO and a helmet manufacturer that most valuations would fall below this threshold. The manufacturer sold the helmets used in this study at a price of Ksh 580, about 15% 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 agent knows that their valuation exceeds the maximum draw. To hedge against the risk of setting the max payout too low, and to avoid anchoring or signaling effects, enumerators did not disclose the upper bound or value of the helmet.[15]

In practice, helmet valuations exceeded the wholesale price of the helmets in about 75% of cases. Figure A2 demonstrates that there is no unusual behavior in the distribution of bids near the

[15]Surveyors were instructed to reveal this information if asked about it, which only happened once.

maximum draw, consistent with accurate valuation data.

2.5 Randomization

Respondents were assigned to the information arms using a pseudo random number drawn IID using SuveyCTO. During wave one, respondents were assigned to the pure control group with probability 0.1 and each of the other groups with probability 0.3. During the second wave, the pure control was eliminated and respondents were assigned to the remaining arms with equal likelihood.

3 Model and identification

This section presents a simple model of demand for motorcycle helmets and introduces the new methodology used for estimating demand for non-market goods, showing how it is used to identify VSL. Furthermore, the model illustrates the assumptions needed to identify VSL under typical approaches assuming rational expectations and demonstrates that these estimates are unbiased only if all individuals have mortality beliefs equal to the econometrician's estimate of risk.

Consider a population of individuals indexed by $i \in \{1, \ldots, \infty\}$. Assume that individuals maximize expected utility and Bayesian update their beliefs about motorcycle risks and helmet effectiveness. Belief formation is modeled to illustrate how a common learning model can generate bias in methods that use empirical risk as a proxy for beliefs. Identification requires a weaker condition that the information presented changes beliefs, which I test empirically.

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

$$Pr(D = 1|\mathcal{I}_i) \sim Beta(a_{iH}, b_{iH}) \tag{1}$$

where $\mathcal{I}_i$ denotes the individual's baseline information set and D is a Bernoulli random variable equal to 1 if the agent dies. The Beta distributed beliefs are over the unknown parameter of the Bernoulli distribution. There is true stochastic variation in whether agents in an accident die, but agents must learn the parameter governing this process. The Beta distribution is natural in this setting because it is a conjugate prior for Bernoulli trials. a_{iH} may be interpreted as the expected number of fatalities and b_{iH} the number of survivals out of $a_{iH} + b_{iH}$ accidents. Define

$$H_{i0} \equiv \mathbb{E}[Pr(D = 1|\mathcal{I}_i)] = \frac{a_{iH}}{a_{iH} + b_{iH}} \tag{2}$$

Suppose that the consumer receives a signal that the estimated efficacy of helmets is $\theta_H \sim Binomial\,(a_{EH} + b_{EH}, a_{EH}/(a_{EH} + b_{EH}))$. The Binomial likelihood captures the fact that studies of helmet efficacy report the number of fatalities out of a given number of accidents. a_{EH} represents the number of fatalities and b_{EH} the number of survivals out of $a_{EH} + b_{EH}$ empirically recorded accidents. The agent's posterior beliefs about the efficacy of helmets are given by

$$Pr(D = 1|\mathcal{I}_i, \theta_H) \sim Beta(a_{iH} + a_{EH}, b_{iH} + b_{EH}) \tag{3}$$

with expected value

$$H_{i1} \equiv \mathbb{E}[Pr(D = 1|\mathcal{I}_i, \theta)] = \frac{a_{iH} + a_{EH}}{a_{iH} + a_{EH} + b_{iH} + b_{EH}} \tag{4}$$

If $\frac{a_{iH}}{a_{iH}+b_{iH}} \neq \frac{a_{EH}}{a_{EH}+b_{EH}}$, the empirical mortality rate differs from the agent's prior expectation, the consumer's posterior mean will differ from their prior ($H_{i0} \neq H_{i1}$). The degree to which beliefs update depends on bias in priors and the precision of their priors and the signal.

The likelihood that a helmet saves one's life is a product of the likelihood of an accident and the effectiveness of a helmet conditional on suffering an accident. Therefore, suppose that the agent has a prior about the probability per trip of getting into an accident that would be fatal to an unhelmeted passenger given by

$$Pr(A = 1|\mathcal{I}_i) \sim Beta(a_{iA}, b_{iA}) \tag{5}$$

where A is a Bernoulli random variable equal to 1 if a trip ends in an accident. Beliefs are again over the unknown parameters of this distribution. Modeling belief formation helps think about what information sources are likely to violate rational expectations. If agents learn through their experiences or their social network then their beliefs will likely vary from empirical estimates. Learning from one's experiences is subject to survivor bias and learning through one's social network is prone to small sample biases (Rabin, 2002). In addition, a higher likelihood of learning about fatal trips, suggested empirically by evidence that bad news travels faster, would cause agents to overestimate risk (Fang and Ben-Miled, 2017).

Suppose the consumer completes n_i motorcycle rides over the lifespan of a helmet. An unhelmeted agent's expectation of getting into a fatal accident over these trips is given by

$$r_{iu} = 1 - \int_{Pr(A=1|\mathcal{I}_i)=0}^{1} [1 - Pr(A = 1|\mathcal{I}_i)]^{n_i} dPr(A = 1|\mathcal{I}_i) \tag{6}$$

The term $[1 - Pr(A = 1|\mathcal{I}_i)]^{n_i}$ is the probability of surviving n_i trips given that the probability of suffering a fatal accident is $Pr(A = 1|\mathcal{I}_i)$ on any given trip. Since beliefs about $Pr(A = 1|\mathcal{I}_i)$ has a non-degenerate distribution, one must integrate over beliefs to obtain equation 6.

I assume that helmeted individuals involved in an accident that would otherwise be fatal are deterred from continuing to use motorcycles.[16] Let z_i indicate whether the agent is exposed to the signal of helmet effectiveness θ. Their subjective probability of suffering a fatal motorcycle accident with a helmet is

$$r_{ih}(z_i) = \begin{cases} H_{i0} \cdot r_{iu}, & z_i = 0 \\ H_{i1} \cdot r_{iu}, & z_i = 1 \end{cases} \tag{7}$$

which is the unhelmeted risk times the perceived chance that a helmet fails to prevent death. The subscript u refers to the agent's perceived unhelmeted risk and h to their helmeted risk. Let p_i denote the price of a helmet. The present value of the agent's utility from being alive is given by $u_a(x_i)$ where x_i is a vector of characteristics, such as income and health. Denote their flow utility of consumption by $u(c_i; x_i)$ and denote the expected utility from not being alive by $u_d(x_i)$.[17] The agent's expected utility from purchasing a helmet is

$$U_{ih}(z_i) = \zeta_h + [1 - r_{ih}(z_i)] \cdot u_a(x_i) - p_i \cdot u'(c_i; x_i) + r_{ih}(z_i) \cdot u_d(x_i) + \epsilon_{ih} \tag{8}$$

[16] This is based on discussions with respondents during the pilot and the fact that people often quit using motorcycles after serious accidents. Absent this assumption, $r_{ih0} = \frac{H_{i0} \cdot r_{in}}{1 - H_{i0} \cdot r_{in}}$. Results are similar using this calculation, differences between the two values are small since the probability of suffering two serious accidents is low.

[17] This may not be zero, for instance if the agent believes in an afterlife.

The parameter ζ_h captures average utility from characteristics of helmets other than safety, and ϵ_{ih} denotes idiosyncratic variation in utility. Without purchasing a helmet, expected utility is

$$U_{iu} = (1 - r_{iu}) \cdot u_a(x_i) + r_{iu} \cdot u_d(x_i) + \epsilon_{iu} \tag{9}$$

Setting $\epsilon_i = \epsilon_{ih} - \epsilon_{iu}$ and $\Delta r_i(z_i) = r_{iu} - r_{ih}(z_i)$,

$$U_{i,h-u}(z_i) \equiv U_{ih}(z_i) - U_{iu} = \zeta_h + \Delta r_i(z_i) \cdot [u_a(x_i) - u_d(x_i)] - p_i \cdot u'(c_i; x_i) + \epsilon_i \tag{10}$$

In words, one's expected utility from a helmet is a function of non-safety preferences for helmets, the agent's *belief* about the probability that the helmet will save their life ($\Delta r_i(z_i)$) times differences in expected utility from surviving versus not, less the price of a helmet times the marginal utility of consumption. The goal is to identify VSL, the change in income needed to compensate an agent for a change in mortality risk. Totally differentiating,

$$dU_{i,h-u} = \frac{\partial U_{i,h-u}}{\partial \Delta r_i} d\Delta r_i + \frac{\partial U_{i,h-u}}{\partial p_i} dp_i \tag{11}$$

Setting $dU_{i,h-u} = 0$,

$$VSL_i \equiv \frac{dp_i}{d\Delta r_i} = \left(\frac{\partial U_{i,h-u}}{\partial \Delta r_i} \right) \bigg/ \left(\frac{\partial U_{i,h-u}}{\partial p_i} \right) = \frac{u_a(x_i) - u_d(x_i)}{u'(c_i; x_i)} \tag{12}$$

This expression predicts that VSL will vary with income. Theory suggests $u_a(x_i)$ will increase and $u'(c_i; x_i)$ will fall as agents become richer. The falling marginal utility of consumption is likely to be particularly important since it would drop linearly with income under log utility and faster under other common utility representations. This reflects the fact that the opportunity cost of funds used to invest in safety is lower when agents are wealthier, so their VSL will be higher even if they obtain no more utility from averting a death.

3.1 Identification under rational expectations

For simplicity, suppose that $\beta = u_a(x_i) - u_d(x_i)$ and $\alpha = u'(c_i; x_i)$ are homogeneous. Denote the econometrician's estimate of the probability that a helmet will save an individual's life by Δr_i^*. A common set of identifying assumptions in the VSL literature (but not this paper) is

$$\text{Full Information Rational Expectations:} \quad \Delta r_i = \Delta r_i^* \quad \forall i \tag{13a}$$

$$\text{Exogeneity and Logit Errors:} \quad \epsilon_{ih}, \epsilon_{iu} \sim_{iid} EV1 \tag{13b}$$

Given data denoting whether agents purchased helmets y_i, empirical risks Δr_i^* and prices p_i,

$$Pr(y_i = 1) = \Lambda\left(\zeta_h + \beta \cdot \Delta r_i^* - \alpha \cdot p_i\right) \tag{14}$$

Λ is the Logistic CDF and so α, β (and $VSL = \frac{\beta}{\alpha}$) may be estimated via maximum likelihood.[18]

Assumption 13a is violated if $Pr(\Delta r_i \neq \Delta r_i^*) > 0$, meaning that some agents have beliefs about risk that differ from the econometrician's estimate, even if Δr_i^* matches beliefs in expectation. The model illustrates that this assumption is only plausible if consumers have the same information as the researcher, which would fail if they possess private information (e.g. experiences or social network reports), or if they cannot access data that the researcher considers. Rational expectations has been assumed despite these limitations because of a lack of an incentive-compatible alternative (prior to this paper) capable of producing meaningful estimates since beliefs are measured noisily.[19] An endogenous relationship between mortality risk and unobserved determinants of utility would violate assumption 13b. In this setting, those that ride motorcycles frequently face greater mortality risk, but they may incur greater disutility from the discomfort of a helmet. In some instances, an instrument is used to relax assumption 13b to a weaker version of rational expectations, but beliefs must still match empirical risk on average (e.g. Ito and Zhang, 2020).

3.2 Identification using subjective belief data

This study estimates VSL without assuming rational expectations or that mortality risk is exogenous by leveraging an instrument that exogenously shifts agents' subjective beliefs. The empirical application shows that one way to generate an instrument is to experimentally update beliefs, but

[18] In practice studies often estimate mixed logit models, but the assumptions are similar.

[19] 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."

one could also follow this approach without experimental variation.[20] The model shows that this is a natural way to identify VSL since presenting agents with information about helmet efficacy will change the expected utility they obtain from the life saving potential of a helmet (unless their priors align with the signal or are degenerate). Updating beliefs about helmet efficacy produces the same identifying variation that exogenous differences in mortality risk would under rational expectations, but this approach is robust to biased beliefs and endogeneity. Formally, assume

$$\text{Exclusion Restriction:} \quad Cov\left(z_i, \epsilon_i\right) = 0 \tag{15a}$$

$$\text{Instrument Relevance:} \quad Cov(\Delta r_i, z_i) \neq 0 \tag{15b}$$

Letting v_i denote willingness to pay for a helmet, an agent's indifference point between paying v_i for a helmet versus not is given by

$$\begin{aligned} \epsilon_{in} &= \zeta_h + \beta \cdot \Delta r_i(z_i) - \alpha \cdot v_i + \epsilon_{ih} \\ \alpha v_i &= \zeta_h + \beta \cdot \Delta r_i(z_i) + \epsilon_i \\ v_i &= \frac{1}{\alpha}\zeta_h + \frac{\beta}{\alpha} \cdot \Delta r_i(z_i) + \frac{1}{\alpha}\epsilon_i = \zeta_h' + VSL \cdot \Delta r_i(z_i) + \epsilon_i' \end{aligned} \tag{16}$$

Hence, VSL is identified from data on v_i and (potentially misreported) beliefs $\Delta r_i(z_i)$, using information assignment z_i to instrument for beliefs (ζ_h' is an intercept). The exclusion restriction is generally satisfied if z_i is randomly assigned unless there is non-classical measurement error, which motivates the focus on testing this condition in this paper. In fact, when one estimates the model using a randomly assigned information treatment as an instrument, then letting Δr_i^t denote the agent's true belief, one may replace assumption 15a with

$$\text{Classical Measurement Error:} \quad \mathbb{E}[\Delta r_i - \Delta r_i^t] = 0 \tag{15a$'$}$$

The model shows that relevance is typically true under Bayesian learning, but a strength of this method is that estimates are consistent as long as information changes beliefs (even if agents learn

[20] I consider a binary instrument for consistency with the model of beliefs, but results are the same with a continuous instrument.

in non-standard ways). Relevance may be verified empirically by testing that $Cov(\Delta r_i, z_i) \neq 0$.[21]

This same framework may be used to estimate demand for other non-market goods by letting r_i be the probability of exposure to the good and z_i be an instrument shifting subjective beliefs about this probability. For instance, when studying demand for clean air r_i may be the probability of exposure to unclean air mitigated by an air purifier.

4 Data and empirical specification

4.1 Data

I use data from 1,571 surveys completed in two waves. The first wave was conducted from October to December 2022, and the second between February and March 2023. The first wave included 921 surveys, counting 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 piloting. 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. Posteriors about the respondent's 5 year mortality risk without a helmet and the effectiveness of helmets at preventing death were collected. Distinct measures were obtained to validate the variables against each other.

Briefly, priors and posteriors about unhelmeted risk were elicited by first having respondents select what range of probabilities their beliefs fell into from a set of bands that span the entire unit interval (such as less than 1 in 10 million, 1 in 10 million to 1 in 1 million, etcetera up to greater than 1 in 10). Then respondents were then asked to list their exact belief within the interval. This two-step approach reduced rounding of low probabilities in piloting.

Table 1 summarizes demographics of the sample and demonstrates balance across treatment arms. Appendix Table A1 is similar, but it examines motorcycle taxi use and prior beliefs. These

[21] If the econometrician also observes data on expected remaining lifespan, LS_i, then the value of a statistical life year (VSLY) is also identified since $v_i = \zeta'_h + VSLY \cdot \Delta r_i(z_i) \cdot LS_i + \epsilon'_i$

outcomes are also generally balanced across experimental groups, although there is some imbalance between the pure control (which had a small number of respondents) and other arms.[22]

4.2 Experimental VSL estimation

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

$$\begin{aligned} v_i &= \zeta_h' + VSL\Delta r_i + X_i'\gamma_0 + \gamma_1 r_{0,i} + \epsilon_i' \\ \Delta r_i &= Z_i'\pi + X_i'\pi_c + \pi_r r_{0,i} + \nu_i \end{aligned} \tag{18}$$

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 lifespan of a helmet, X_i is a vector of controls, and $r_{0,i}$ is the respondent's baseline belief about their unhelmeted mortality risk. I report results with two sets of excluded instruments, Z_i. First, I consider an "interacted" set $Z_i = (T_i', r_{0,i} \cdot T_i')'$ where T_i is a vector of treatment assignment indicators. The second set consists of treatment only, $Z_i = T_i$.

The preferred estimate uses the interacted instruments because they absorbs heterogeneity in priors, improving power. Intuitively, if there is variation in the perceived riskiness of motorcycles, then Δr_i will vary both due to beliefs about motorcycle risks and the efficacy of helmets. The interacted instruments model this fact, so the first stage will more accurately capture posterior beliefs. This set of instruments is similar to that studied in Abadie et al. (2023), which shows that the interacted first stage yields improved asymptotic mean squared error.[23]

Controls are selected using single-post LASSO. Possible controls include demographic variables, motorcycle trip characteristics, and the information sources used to construct beliefs about mortality risk. Estimates also include surveyor fixed effects.[24]

I follow the pre-analysis plan 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 to help rule out confounds such as demand effects. The primary tables report homoskedastic standard errors

[22] Appendix Table A2 and Appendix Table A3 demonstrate that non-response rates are also similar across treatment arms and that response rates are high.

[23] I pre-specified two sets of instruments before the first wave of the experiment. The original pre-analysis plan (PAP) specified the use of n_i, the number of motorcycle trips taken per week, in place of $r_{0,i}$. This assumed n_i would effectively proxy for $r_{0,i}$, which is not the case. A PAP amendment filed before wave 2 specified the use of $r_{0,i}$.

[24] 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 and enumerator FEs. Results are similar if they are excluded.

and weak instrument robust confidence sets. Results are similar using GMM with heteroskedastic robust errors. Appendix B provides details, including justifications for these decisions.

4.3 Alternative VSL estimation procedures for comparison

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

$$v_i = \zeta_h + VSL\Delta r_i^* + X_i'\gamma_0 + \epsilon_i \tag{19}$$

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 random coefficient logit models using empirical risk estimates as a proxy for beliefs. I similarly estimate a demand model using empirical estimates of mortality risk, but the specification is linear since willingness to pay is observed. Following León and Miguel (2017), I do not instrument for Δr_i^*.

Second, I report the estimates

$$VSL_i = \frac{v_i}{\Delta r_i^*} \tag{20}$$

This approach assumes that agents do not receive utility from any characteristics of helmets other than their life saving potential and that $\Delta r_i = \Delta r_i^*$. This specification is identical to Berry et al. (2020) and is similar to Kremer et al. (2011). The estimator is applied in the context of safety products, so the assumption that agents only consider the safety benefits of products is plausible. This approach allows one to estimate VSL in settings where cross-sectional variation in mortality risk is small or unobserved. Neither approach separates willingness to pay for non-fatal illness or injury prevention from the agents' demand for mortality risk reduction, so authors often instruct readers to interpret estimates as upper bounds (León and Miguel, 2017; Kremer et al., 2011).

5 Results

5.1 Estimates of the value of a statistical life

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, supporting the use of subjective belief data to estimate VSL.

The mean baseline belief about one's 5 year mortality risk is 0.034, the median is .001 and the 99th percentile is 0.5.[25] For comparison, I estimated that the median empirical risk was about 1 in 4,000, but the data may underestimate deaths because the NTSA only record fatalities in which a victim is dead on arrival (Muguro et al., 2020). 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 is consistent with overoptimism in one's own safety.

Beliefs about the 5 year and 1 year mortality rates of motorcycle passengers are also 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 (less than 1%) stated that they believed more people would die over the next 1 year than the next 5 years, which is impossible.[26] 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 ($p < .01, R^2 = 0.2$).

First stage: Effect of information on beliefs

Table 2 demonstrates that randomized information exposure had a statistically significant effect on the agent's posterior belief that a motorcycle helmet will save their life.

I first presented respondents with estimates of their unhelmeted mortality risk constructed using data from the NTSA, which has a low level of trust. Consistent with mistrust, respondents reported no change in beliefs in response to this information as shown in columns 1 and 2. There is not a significant difference between either treatment arm, which received identical information, and the control group. This provides reassuring evidence that agents did not feel compelled by

[25] VSL estimates are similar if implausible responses are excluded or beliefs are winsorized.
[26] Estimates are similar if these observations are dropped.

experimenter demand effects to report changes in posteriors when no updating occurred.

Respondents did update their beliefs about the effectiveness of helmets based on the studies presented in the treatments, which is the primary variation used by the interacted instruments. Respondents *overestimated* the effectiveness of helmets at baseline and updated their beliefs downward when exposed to information. Table 2 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%, and many respondents stated beliefs over 90%. The average in the high treatment group was about 75%, and that in the low treatment group was about 65%. The differences between the treatments arms are significant at the 1% level.

The change in beliefs about helmet effectiveness is sufficient to detect a change in the probability that a helmet would save the respondent's life (the product of the prior two variables). The dependent variable in Columns 5 - 6 of Table 2 equals 10,000 times the respondent's posterior mean. This value is about 80 units (36%) lower on average in the low treatment versus control group. The low treatment beliefs 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 (because the high treatment information is similar to priors), but point estimates change in the expected direction. Overall, one can reject the null hypothesis that the information treatments had no effect on beliefs, meaning that treatment assignment is a relevant instrument for the perceived mortality risk reduction offered by a helmet.

The value of a statistical life

I report estimates of the value of a statistical life, constructed using the new experimental method, in Table 3. The results suggest that demand for safety is low in this sample: the preferred estimate of VSL reported in column 3 is just USD PPP $224, and a weak instrument robust confidence set rejects values below $34 and above $429 with 95% confidence.[27] Results are similar with and without controls and using the "interacted" or "treatment only" instrument sets.

The estimates pass a battery of robustness checks. Panel b shows that excluding control observations does not significantly change results, and one may reject values above $1,038 with 95% confidence despite the loss in power due to a smaller sample size. Panel c similarly shows that leaving any experimental arm out does not meaningfully affect the results.

[27] This is the preferred estimate because it uses the interacted instrument set which is better powered.

This method for estimating VSL does not require experimental variation if a valid instrument for subjective beliefs exists in the observational data. Leveraging this fact, I instrument for risk using a plausibly exogenous shifter of beliefs in column 5 of panels a and b: an indicator for whether the respondent knows a motorcycle accident victim, controlling for taxi terminal fixed effects. Estimates are similar to those obtained from random variation. Column 6 uses victim exposure and treatment assignment as instruments, and a J-test of over-identifying restrictions fails to reject the equality of estimates across variation sources.[28]

Table 3 excludes 33 observations that were contaminated by motorcycle drivers pretending to be passengers, submitting false survey responses to receive a free helmet.[29] Appendix Table A7 reports VSL estimates with the manipulated surveys included. Results do not substantially change, although confidence sets are wider. Appendix Table A9 demonstrates that results also hold if beliefs about the probability that a helmet will save one's life are winsorized. Appendix Table A8 shows that the results are robust to changes in planned future ridership due to receiving a helmet (panel a) or subjective beliefs about the lifespan of a helmet (panel b). Panel c suggests that results do not disproportionately weight individuals that select into high motorcycle ridership because of low demand for safety.

Although demand for safety is low in this sample, VSL estimates are consistent with theory and much higher estimates among wealthy populations. As discussed in section 3, theory predicts that VSL estimates should be higher among high-income populations because the opportunity cost of funds falls with wealth. Comparing results from this paper to revealed preference estimates across contexts reveals that the values reported in this paper are of a similar order of magnitude to those from similar income levels, but they are much smaller than those from high-income populations.

Figure 1 plots point estimates and standard errors from all revealed preference studies of VSL in LMICs that I am aware of.[30] In addition, I include an estimate from Greenberg et al. (2021), which examines a population of US soldiers, for comparison to a high-income setting using similar

[28]This comparison uses the "treatment only" instruments because controlling for priors would invalidate the exposure instrument. Columns 5 and 6 report heteroskedastic-robust standard errors since the use of homoskedastic errors in other cases is based on the random assignment of treatments.

[29]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.

[30]This includes Kremer et al. (2011), León and Miguel (2017), Berry et al. (2020), Ito and Zhang (2020), and Shrestha (2020). I convert estimates to 2022 dollars using the change in the CPI from the paper's publication year. I plot estimates using a scale that is linear near zero but logarithmic far from zero and truncated below -\$1.

methods. Point estimates presented in this paper are lower than most prior values, but similar to those from populations with similar incomes. For instance, Kremer et al. (2011) estimate a value of about USD $1,000 to avert a child death in rural Kenya, and Berry et al. (2020) estimate median VSLs below $0 and just over $4,000 in Ghana. Estimates orders of magnitude larger from Ito and Zhang (2020) and León and Miguel (2017) examine populations with higher incomes, close to $10,000 and $75,000 respectively.

As a thought experiment, one may assume that one's utility from being alive is constant across income levels and calculate what curvature of marginal utility of consumption would be required to rationalize the differences in VSL observed across income levels. This is an upper bound on the true curvature implied if quality of life increases with income. Assuming a CRRA utility function, the differences between the preferred estimate of VSL and those reported in León and Miguel (2017) imply a coefficient of relative risk aversion of $\theta < 3$.[31] These values fall within admissible estimates of θ. In fact, Havránek (2015) find a mean elasticity of inter-temporal substitution of $1/3$ in a meta-analysis, implying $\theta = 3$ under power utility.

Heterogeneity in the value of a statistical life

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 of urban commuters, a sample of interest for transportation policy. However, I examine the degree to which VSL is heterogeneous across 8 different variables in Table 4 to provide insight into the extent to which the average VSL over all residents of Nairobi may differ from these estimates. I find evidence of heterogeneity that is consistent with economic theory, but estimates are sufficiently small that it is unlikely to change the conclusion that willingness to pay for safety is low.

In column 1, I examine heterogeneity by age. Since a younger respondent has more life years remaining in expectation, age would be negatively correlated with VSL if the value of a statistical life year were homogeneous. However, I find that VSL is the same among respondents below and above the median age of 32. These results are consistent with Aldy and Viscusi (2007) which shows that VSL peaks around age 40.

[31] Assume $\beta_i = \beta, \alpha_i = Y_i^{-\theta}$ where Y_i is income. Then $\frac{VSL_1}{VSL_2} = \left(\frac{Y_1}{Y_2}\right)^\theta \Rightarrow \theta = \log\left(\frac{VSL_1}{VSL_2}\right) / \log\left(\frac{Y_1}{Y_2}\right)$. Plugging in the estimates from the two studies, $\theta \approx \frac{\log(700{,}000/224)}{\log(75{,}000/4{,}750)} = 2.91$

I find an economically and statistically significant relationship between VSL and income as shown in columns 2 and 3. The second parameter in column 2 reports an estimate of the interaction between the coefficient on the mortality risk reduction of a helmet and asinh(wages), demeaned so that the first coefficient captures average VSL. The point estimates indicate that a 1% increase in wages in the sample is associated with an increase in VSL of about $4, implying an income-elasticity of VSL of about 1.14. This slightly exceeds cross-country estimates (Masterman and Kip Viscusi, 2018). Column 3 shows agents with above median wages have a VSL about $500 higher.

Column 4 provides evidence that VSL is also increasing with health. There is not significant heterogeneity with respect to having children, performance on a digit span recall test, years of education, or gender.

Welfare implications of VSL estimates

The welfare implications of the VSL estimates presented in this paper depend on the values that are currently used in benefit-cost analysis. If larger VSL values are considered, then switching to the estimates from this paper could yield welfare gains by aligning policy with preferences.

I selected 5 recently published benefit-cost analyses in Kenya to examine how these VSL estimates affect policy conclusions.[32] The 5 analyses selected, which come from 4 different studies, are the 5 most recent identified. I examine how this paper's VSL estimates change the policy conclusions of these studies by replacing the original VSL (or VSLY) estimates used to value health benefits with the preferred estimates from this paper.

Figure A4 shows that benefit-cost ratios (BCRs) drop substantially. BCRs fall by over 99% on average, and in 4 out of 5 instances the ratio falls from a value above to 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.

The results of this paper also indicate that development assistance donations may be misallo-

[32] I identified relevant studies by searching for benefit-cost analyses on Google Scholar that referenced Kenya and utilized VSL (or VSLY). Babagoli et al. (2022) report two analyses associated with providing menstrual cups or sanitary pads to young women, Mwai et al. (2023) examine primary health care investment, Hamze et al. (2017) study a cleft lip and palate repair program, and Oyugi et al. (2023) consider free maternity care.

cated. For instance, GiveWell, an NGO that matches donor funds to charities that has directed over $1 billion, weights averting the death of someone aged 15-49 104 times doubling their consumption for a year.[33] The VSL estimated in this paper indicates that consumers would value a doubling of their annual income more than 10 times as highly as averting a death on average, a three order of magnitude difference. GiveWell uses these weights to calculate the cost effectiveness of programs, informing which are recommended and eligible for certain funds.[34] This paper suggests that weighting consumption more strongly could yield substantial welfare gains.

When does theory support the use of LMIC VSL estimates?

The value of a statistical life is the utility that a consumer receives from a reduction in mortality risk normalized by their marginal utility of consumption. This converts utils, which are uninterpretable, to dollars, allowing one to compare the benefits of safety improvements to costs. But estimates are only valid for decisions where a fixed population faces a mortality risk-consumption trade-off. This applies to important policy decisions and aid allocation like the cases described above. But it does not apply in cases where the funding source of a health program is fixed.

To illustrate this point, suppose that there are homogeneous high-income consumers indexed by h and low-income consumers indexed by ℓ. As in section 3, let α denote the marginal utility of consumption and β one's expected utility from being alive versus not. If the high-income government distributed N vaccines to low-income consumers that save each individual's life with probability p at a cost of $c \cdot N$ paid for via taxes on N high-income consumers, then the global welfare change, in utils, would be

$$W = N \cdot p \cdot \beta_\ell - N \cdot c \cdot \alpha_h \tag{21}$$

Which is positive if and only if $N \cdot p \cdot \frac{\alpha_\ell}{\alpha_h} \cdot VSL_\ell > N \cdot c$. Therefore policy decisions made by estimating lives saved, valuing them at the low-income country's VSL, then comparing these benefits to the program's costs will not maximize welfare. Benefits would be under-estimated by $\frac{\alpha_\ell}{\alpha_h}$, which this paper suggests may be orders of magnitude. Intuitively, low-income consumers

[33] Based on the 2023 update on GiveWell's moral weights.

[34] The weights are based on a stated preference VSL estimate of over USD $38,000 from Kenya (Redfern et al., 2019). Walker et al. (2023) also estimate willingness to pay per DALY at USD PPP 3,611 using a stated preference approach, but note that a revealed preference estimate from the same sample derived from Kremer et al. (2011) produces an estimate of USD PPP 67.

have lower VSLs largely due to a higher opportunity cost of funds, but if high-income consumers pay program costs, low-income consumers face no consumption loss.[35]

One example where this distinction may be important is vaccine policy. For instance, Ahuja et al. (2021) examine how changes in vaccine procurement could reduce deaths, valuing lives saved by VSLs adjusted to countries' income levels. Agrawal et al. (2023) argue that more equitable COVID-19 vaccine allocation may have saved 670,000 lives but reduced welfare by $6.5 trillion because of VSL differences across income levels. But rich countries tend to disproportionately fund vaccine research and most LMICs relied on donated COVID-19 vaccines (Ahuja et al., 2021). Hence, these studies likely underestimate the welfare gains of lives saved in poor countries.

5.2 Are VSL estimates biased by non-classical measurement error?

The primary identifying assumption in this paper is that beliefs are measured with classical error. Tests for mismeasured beliefs built into the experiment and moments of the data indicate that measurement error is likely classical, implying that VSL estimates are consistent.

First, the experiment randomized whether agents received no information, the low signal, or the high signal so that VSL may be estimated at different points of the belief distribution. Estimates should be invariant to the experimental arms included under classical measurement error. But they would often differ if beliefs were misreported, unless agents misstated changes in beliefs by the same proportion in the low and high treatment arms.[36] For instance, if the reduction in perceived helmet efficacy reported in the "low treatment" arm were a product of experimenter demand effects, estimates using the low treatment and control observations would be smaller than those based using high treatment and control observations. Panel c of Table 3 shows that results are similar using any combination of experimental arms.

The treatment only and interacted instrument sets facilitate a second test for mismeasured beliefs. Appendix C, using an argument similar to those in the literature on local average treatment effects (e.g. see Abadie et al., 2023; Mogstad et al., 2018, 2021), shows that the "treatment only" IV weights the VSL of individuals linearly in priors about the riskiness of motorcycles, $r_{0,i}$. The "interacted" instruments weight individuals proportionally to $r_{0,i}^2$. Any bias in posteriors corre-

[35]If high-income consumers are choosing between donations to health or consumption programs as in the case of GiveWell, then there is an LMIC health-consumption trade-off so applying the LMIC VSL is appropriate.

[36]Specifically, respondents would need to report a constant c times their true change in beliefs about helmet efficacy.

lated with bias in priors should therefore cause the instrument sets to produce different results. The appendix shows that common patterns of misreporting – including rounding small probabilities, "S-shaped" probability weighting (Kahneman and Tversky, 1979), and experimenter demand effects – will typically feature such a correlation and be detected by this test. Estimates are similar across the instruments and a Hansen J-test fails to reject their equality ($p > 0.5$), suggesting that significant bias is unlikely.[37]

Several features of the data further support the consistency of estimates. If agents misreported beliefs due to demand effects, then one would expect reported changes in posteriors about the riskiness of motorcycles and effectiveness of helmets since information about both parameters was provided. But there was no reported updating in beliefs about motorcycle risks. Respondents also rarely reported posteriors about helmet effectiveness that exactly aligned with study signals, and helmet demand is lower in the "low treatment" arm by a significant margin, which provides revealed preference evidence of a true change in beliefs. Willingness to pay for a helmet is also similar in the control and pure control groups despite the fact that those assigned to control were asked multiple questions that highlighted the risks of motorcycles and benefits of helmets. Lastly, VSL estimates that leverage quasi-exogenous variation from accident victim exposure, which was measured before treatment delivery, yields statistically indistinguishable estimates.

Results are also inconsistent with systematic over-reporting of beliefs due to cognitive challenges or low education biasing results. Table 4 shows that there is not statistically significant heterogeneity with respect to performance on a digit span recall test or years of education. And VSL estimates are similar when beliefs are winsorized (Appendix Table A9), whereas one would expect VSL estimates to increase if a subset of individuals substantially over-reported risk.

5.3 Testing rational expectations

The primary assumption relaxed by this paper is that agents beliefs satisfy rational expectations. Results suggest that this condition is violated in this context and that VSL estimates obtained using typical methods are biased.

This study captured agents' subjective beliefs and their empirical mortality risk, allowing

[37] This p-value is based on the preferred specification, which also includes multiple treatments. If the high treatment and control groups are pooled so the J-test only compares the interacted and non-interacted instruments the test also fails to reject.

for direct tests of rational expectations. The variables do not have a significant correlation ($\rho = -0.0008$) and respondents report beliefs about 100 times empirical risk on average, suggesting that rational expectations is violated. A weak correlation could be a product of measurement error. One way to correct for this is to instrument for the final belief elicitation using a prior measure from the survey, which eliminates bias from classical measurement error if it is independent across questions (Gillen et al., 2019). This approach does yield a statistically significant positive relationship between the variables, but the estimated coefficient is under 0.001, and I can reject the null hypothesis that it equals 1 (i.e. rational expectations) with confidence greater than 99.9%. However, since this test is sensitive to correlated errors across belief measures, I turn to two other sources of data to examine how agents form expectations.

First, the survey directly asked respondents to list each of the information sources they used to form their beliefs. Figure 2 plots the responses. The most common source cited is own experience (79%), followed by family members (48%) and social media (38%). Under a third of respondents reported that media or government sources, which are more objective, informed their views about safety. These information sources are consistent with deviations from rational expectations since they depend on private information and are prone to bias.

Second, I regress subjective beliefs on variables that are likely to affect mortality risk empirically and given the information sources reported by respondents. I consider trip volume and trip duration as empirical shifters since those that ride motorcycles more often are mechanically more likely to die on one over a fixed time horizon, particularly since respondents do not operate them and I may restrict the comparison to individuals with trips at the same origin and driven by the same operators by including taxi terminal fixed effects. Validating this empirically, I find that a 1 percent increase in reported motorcycle taxi trips in a typical week is associated with about a 0.13 percentage point increase in the probability of reporting a prior accident ($t \approx 5$). I consider indicators equal to 1 if the respondent was in any prior motorcycle accident or if they have a social network contact that experienced a motorcycle accident to test agents' reports that they predominately consider the experiences of themselves or their associates when forming beliefs.

If rational expectations were satisfied, then one would expect positive coefficients on the empirical shifters of risk. Table 5 demonstrates that this is not the case: beliefs about risk are unrelated to trip volume or the length of a typical trip with or without terminal fixed effects and covariates selected via double-post LASSO (Belloni et al., 2014). In contrast, the accident history of the respondent or their social network is strongly positively correlated with beliefs, with or without terminal fixed effects and covariates, consistent with respondents' reports. This suggests that the lack of a relationship between beliefs and empirical determinants of risk is unlikely a product of measurement error and that rational expectations is unlikely to hold.

One concern with these results is that the prior accident history may be driven by risk heterogeneity observed to the respondent but not the econometrician, in which case one may conclude that rational expectations holds with a more accurate model of empirical risk. Even if this were true, the findings demonstrate the sensitivity of results to the data that the researcher uses. However, to further test this hypothesis, I examine whether respondents reported higher 5-year risks if it rained the day of the survey. The weather today should not affect long-run risk empirically, but rain increases the salience of conditions when motorcycles are dangerous (Bordalo et al., 2013). Respondents reported about twice the risk if it rained ($p < .01$) and covariates and terminal fixed effects have almost no effect on this estimate, suggesting a departure from rational expectations.

I next examine if the approaches typically used to estimate VSL, which assume rational expectations, produce biased estimates of VSL in Table 6.[38] I find that estimates generally do not fall in experimental confidence sets. In addition, different observational approaches produce starkly different estimates of demand for safety.

Columns 1 - 4 report estimates of VSL obtained from Equation 19. The first column does not include any covariates. Columns 2 - 4 are estimated using double-post LASSO to select controls. Column 3 adds enumerator fixed effects and column 4 includes taxi terminal fixed effects. Estimates are sensitive to the covariates and fixed effects included, and confidence intervals are wide. The point estimate is over $15,000 without controls and is negative in column 3. Columns 2 and 4 produce point estimates that are close to experimental estimates, about $2,700 and just under $1,500 respectively. This provides the most promising evidence that demand estimates that assume rational expectations can consistently estimate VSL. However, standard errors are large and these results may be driven by a weak correlation between willingness to pay and ridership, rather than

[38] In order to calculate the empirical likelihood that a helmet will save an individual's life, the econometrician must take a stance of the efficacy of helmets. Panel a reports results assuming that the Liu et al. (2008) estimate of 42% is the truth, and panel b reports estimates using 70% efficacy as estimated in Ouellet and Kasantikul (2006). Discussion focuses on panel a.

the approach accurately estimating demand for safety.

An alternative approach to estimate VSL from observational data that can produce precise estimates even if the parameter is low is given by Equation 20. Berry et al. (2020) use this approach with elicited willingness to pay data, and Kremer et al. (2011) report estimates that depend on similar assumptions. I report estimates using this approach in column 5. The mean estimate is over \$380,000, and the confidence interval excludes values below \$354,000.

The experimental methodology relaxes rational expectations and ensures that endogeneity does not affect estimates. A natural question is the extent to which each of these forces drives the disparity between estimates. Column 6 therefore estimates VSL using subjective belief data, instrumenting for the final measure of beliefs using their prior about the number of motorcycle deaths in Nairobi over 5 years per 10,000 passengers. If measurement error is independent across belief measures, this overcomes attenuation bias, but it will not purge estimates of endogenous belief formation (Gillen et al., 2019). This yields an estimate of about \$110, which is smaller than experimental point estimates but not different by a statistically significant margin. While suggestive, this indicates that deviations from rational expectations may be more binding than endogeneity in this setting. This is consistent with the interpretation that subjective beliefs are heavily influenced by idiosyncratic factors, such as accident realizations, which would lead to deviations from rational expectations but limit endogeneity.

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 estimate demand for motorcycle helmets in East Africa. Helmet producers and road safety organizations attributed low helmet use to a lack of demand prior to this study. This study is not designed to evaluate programs to improve helmet use, but it can test whether demand is low. 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 7 indicates that helmet demand is not deflated by consumers underestimating the safety benefits of helmets. The effects of the treatment are small on average, and presenting consumers with information reduces their willingness to pay for helmets.

Although use of helmets is rare in the setting, Figure 3 suggests that demand for helmets may not be low. The figure plots inverse demand and the elasticity of demand among control respondents. [39] The wholesale price of helmets is denoted by a vertical red line. I estimate that over 75% of motorcycle taxi passengers were willing to forego a cash payment at least as large as the wholesale cost of helmets to receive one, and about 60% were willing to forego Ksh 1,000 or more, which is over 170% of the wholesale cost. This suggests that there may be unmet demand for helmets, although relaxing liquidity constraints when eliciting demand could drive this conclusion.

High demand is reconciled with low VSL estimates by two facts. First, as outlined in section 5.1, consumers view motorcycles as quite dangerous. So the change in willingness to pay induced by the information interventions is small relative to the change in 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. 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 are costly due to lost wages and medical bills.

Further supporting the view of unmet demand, the helmet manufacturer reported easily selling helmets when they sent vans to motorcycle taxi stands. This suggests that the retail market for helmets may be thin. Qualitative evidence supports this conclusion. Many respondents reported that they believed that helmets were only sold it a bundle with motorcycles. The low retail availability of helmets may be driven in part by the fact that the manufacturer had only been producing helmets for about a year at the time of the study.

7 Conclusion

This study introduces a framework for experimentally estimating demand for non-market amenities using subjective belief data and validates the approach by estimating the value of a statistical life. I leverage the fact that products often exist which affect an agent's exposure to such goods. This paper demonstrates that researchers may update agents' beliefs about how a product affects the amenity of interest, then examine how product demand changes. This method is tractable and low-cost in many settings where exogenous variation in attribute exposure or product characteristics

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

does not exist, and it avoids the need to assume that agents' beliefs satisfy rational expectations.

I estimate low demand for mortality risk reduction in Kenya, consistent with the lower tail of estimated East African VSLs but far below values used in policy. These results suggest that there is room to improve welfare by aligning the decisions of governments and NGOs with preferences. Highlighting potential misallocation, the conclusions of 4 out of 5 benefit-cost analyses that I examined change when the VSL estimates from this paper are applied, and I show that a prominent aid organization under-weights consumption gains by orders of magnitude compared to the efficient level estimated in this paper. I find a VSL with respect to income exceeding 1, illustrating the importance of applying context-specific estimates.

This paper also provides evidence that agents beliefs violate rational expectations, a core assumption in most other VSL estimates. Belief data supports agents' reports that they consider information from personal experiences, which are prone to selection and survivor bias, to form expectations. Consistent with bias, agents' beliefs are orthogonal to empirical shifters of mortality risk. This finding helps rationalize the significant dispersion of VSL estimates that assume rational expectations and demonstrates the value of the tools introduced in this paper which facilitate the precise estimation of the parameter from subjective belief data. This framework may also be useful for estimating demand for other non-market goods where consumers' beliefs may be biased, such as environmental amenities (Baylis et al., 2023).

References

Abadie, A., J. Gu, and S. Shen (2023, 3). Instrumental variable estimation with first-stage heterogeneity. *Journal of Econometrics*, 105425.

Agrawal, V., N. Sood, and C. M. Whaley (2023, 10). The Impact of the Global COVID-19 Vaccination Campaign on All-Cause Mortality. *NBER Working Paper*.

Ahuja, A., S. Athey, A. Baker, E. Budish, J. C. Castillo, R. Glennerster, S. D. Kominers, M. Kremer, J. Lee, C. Prendergast, C. M. Snyder, A. Tabarrok, B. J. Tan, and W. Wiecek (2021, 5). Preparing for a Pandemic: Accelerating Vaccine Availability. *AEA Papers and Proceedings 111*, 331–35.

Aldy, J. E. and W. K. Viscusi (2007, 7). Age Differences in the Value of Statistical Life: Revealed Preference Evidence. *https://doi.org/10.1093/reep/rem014 1*(2), 241–260.

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.

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.

Baylis, P., M. Greenstone, K. Lee, and H. Sahai (2023). Is the Demand for Clean Air Too Low? Experimental Evidence from Delhi.

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.

Blass, A. A., S. Lach, and C. F. Manski (2010, 5). USING ELICITED CHOICE PROBABILITIES TO ESTIMATE RANDOM UTILITY MODELS: PREFERENCES FOR ELECTRICITY RELIABILITY*. *International Economic Review 51*(2), 421–440.

Bordalo, P., N. Gennaioli, and A. Shleifer (2013, 10). Salience and consumer choice. *Journal of Political Economy 121*(5), 803–843.

Campbell, H. F. and R. P. C. Brown (2003, 7). Valuation of Non-marketed Goods. *Benefit-Cost Analysis*, 261–287.

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.

Dupas, P. and E. Miguel (2017, 1). Impacts and Determinants of Health Levels in Low-Income Countries. *2*, 3–93.

Fang, A. and Z. Ben-Miled (2017, 3). Does bad news spread faster? *2017 International Conference on Computing, Networking and Communications, ICNC 2017*, 793–797.

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*.

Gillen, B., E. Snowberg, and L. Yariv (2019, 8). Experimenting with measurement error: Techniques with applications to the caltech cohort study. *Journal of Political Economy 127*(4), 1826–1863.

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.

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.

Havránek, T. (2015, 12). Measuring Intertemporal Substitution: The Importance of Method Choices and Selective Reporting. *Journal of the European Economic Association 13*(6), 1180–1204.

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.

Kahneman, D. and A. Tversky (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica 47*(2), 263–292.

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.

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).

Manski, C. F. (2004, 9). Measuring Expectations. *Econometrica 72*(5), 1329–1376.

Masterman, C. J. and W. Kip Viscusi (2018, 9). The Income Elasticity of Global Values of a Statistical Life: Stated Preference Evidence. *Journal of Benefit-Cost Analysis 9*(3), 407–434.

Mendelsohn, R. and S. Olmstead (2009, 11). The economic valuation of environmental amenities and disamenities: Methods and applications. *Annual Review of Environment and Resources 34*(Volume 34, 2009), 325–347.

Mogstad, M., A. Santos, and A. Torgovitsky (2018, 9). Using Instrumental Variables for Inference About Policy Relevant Treatment Parameters. *Econometrica 86*(5), 1589–1619.

Mogstad, M., A. Torgovitsky, and C. R. Walters (2021, 11). The Causal Interpretation of Two-Stage Least Squares with Multiple Instrumental Variables. *American Economic Review 111*(11), 3663–98.

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

Muguro, J. K., M. Sasaki, K. Matsushita, and W. Njeri (2020, 1). Trend analysis and fatality causes in Kenyan roads: A review of road traffic accident data between 2015 and 2020. *http://www.editorialmanager.com/cogenteng 7*(1), 1797981.

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.

Nyachieo, G., V. Wandera, M. Peden, and S. Clark (2024). A Fare Price: An Investigation into the Health Costs of Motorcycle Taxi Crashes in Kenya. Technical report, Transaid.

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.

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.

Tatah, L., L. Foley, T. Oni, M. Pearce, C. Lwanga, V. Were, F. Assah, Y. Wasnyo, E. Mogo, G. Okello, S. Mogere, C. Obonyo, and J. Woodcock (2023, 6). Comparing travel behaviour characteristics and correlates between large and small Kenyan cities (Nairobi versus Kisumu). *Journal of Transport Geography 110*, 103625.

Walker, M. W., A. H. Huang, S. Asman, S. J. Baird, L. Fernald, J. Hamory, H. Fernando, H. De La Guardia, S. Koiso, M. Kremer, M. N. Krupoff, M. Layvant, E. Ochieng, P. Suri, and E. Miguel (2023, 4). Intergenerational Child Mortality Impacts of Deworming: Experimental Evidence from Two Decades of the Kenya Life Panel Survey.

Wiswall, M. and B. Zafar (2018, 2). Preference for the Workplace, Investment in Human Capital, and Gender. *The Quarterly Journal of Economics 133*(1), 457–507.

Figure 1: Revealed-preference VSL estimates across studies

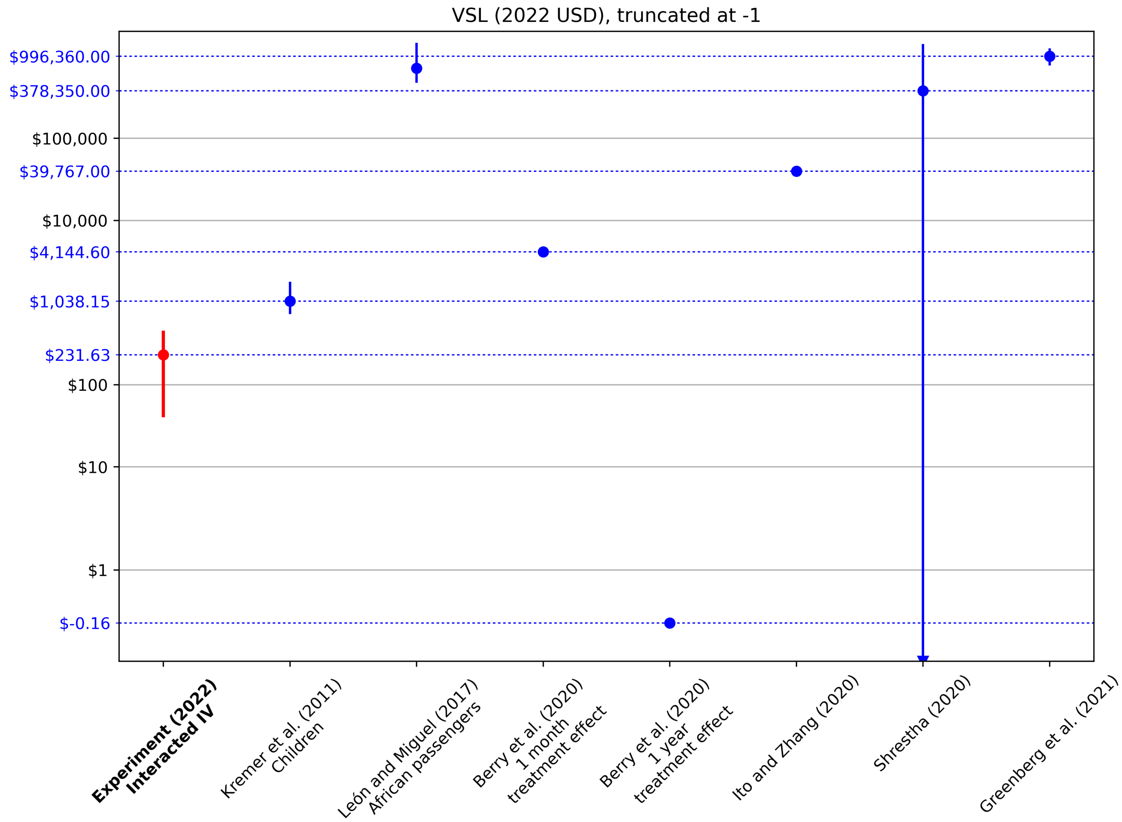


Figure 1 plots revealed-preference VSL estimates and (where available) 95% confidence intervals from this paper, Kremer et al. (2011), León and Miguel (2017), Berry et al. (2020), Ito and Zhang (2020), Shrestha (2020), and Greenberg et al. (2021). Greenberg et al. (2021) is included for comparison to a high-income setting. The other estimates are from low and middle income economies. All estimates are presented in 2022 USD calculated by inflating based on the paper's publication year using the CPI inflation calculator. The lower bound of the 95% confidence interval from Shrestha (2020), which is below -$600,000, is truncated at -$1. Ito and Zhang (2020) does not report a confidence interval. Appendix Figure A3 plots a version of the figure without truncation.

Figure 2: Information sources used to form beliefs about motorcycle safety

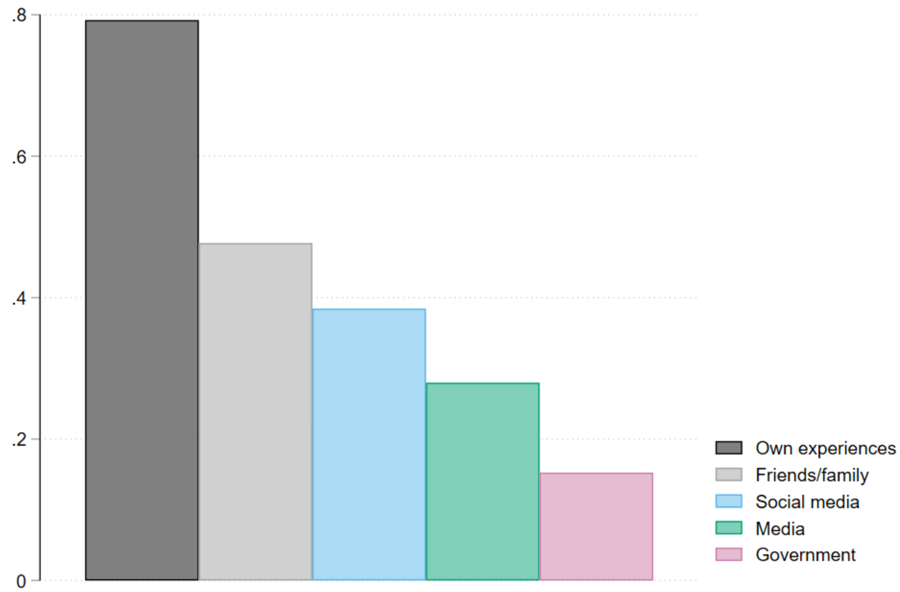
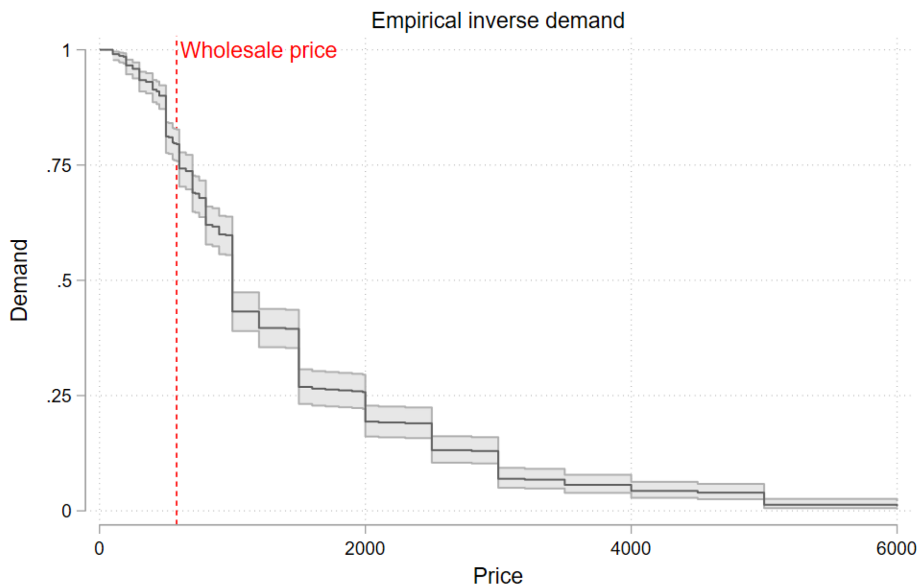


Figure 2 plots the information sources that respondents reported using to form beliefs about the mortality risk of motorcycles. Respondents were able to select multiple options, so the columns do not add to 1.

Figure 3: Demand for helmets

(a) Inverse demand, control group



(b) Elasticity of demand, control group

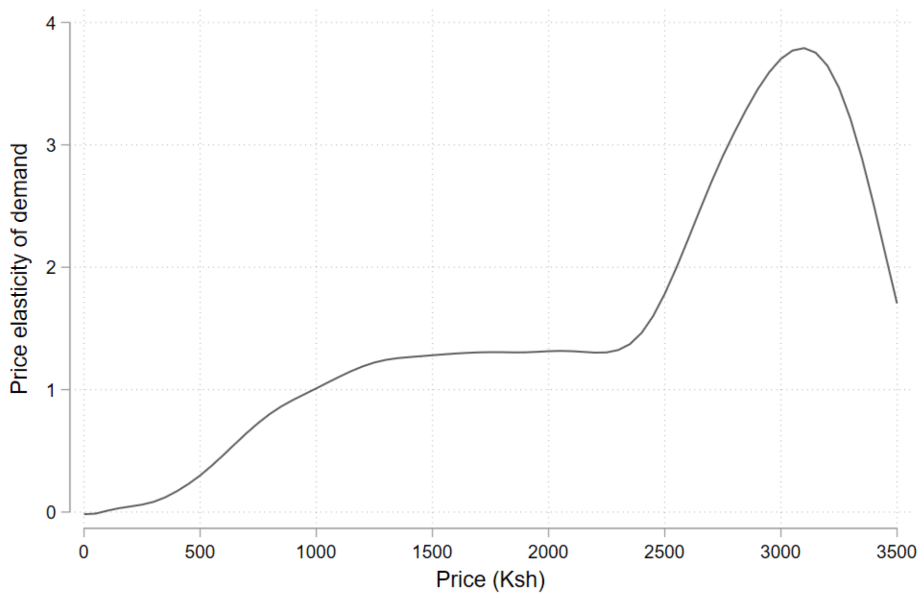


Figure 3 plots estimates of demand for helmets among observations in the control and pure control arms. The plot of demand includes a pointwise confidence interval. The plot of the elasticity of demand is based on a local polynomial estimation adapted from Berry et al. (2020). The vertical line denotes 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.

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	0.282	0.659	0.377
	[8.126]	(1.159)	(0.572)	(0.589)	(0.590)
Female	0.352	-0.039	-0.064**	-0.050	0.014
	[0.478]	(0.060)	(0.030)	(0.030)	(0.029)
Health (1-5)	3.403	-0.034	-0.056	-0.012	0.044
	[0.667]	(0.086)	(0.043)	(0.044)	(0.043)
Life expectancy	81.438	-0.134	0.457	-0.119	-0.577
	[6.947]	(0.840)	(0.427)	(0.439)	(0.413)
Employed	0.898	-0.051	-0.041*	-0.047**	-0.006
	[0.303]	(0.043)	(0.022)	(0.022)	(0.022)
Income (PPP, ’000s USD)	6.866	-0.150	0.635	1.017	0.384
	[8.427]	(1.472)	(0.834)	(0.862)	(0.956)
$\mathbb{E}$[Wage in 5 years/Wage today]	6.166	-0.636	-0.925	-1.422*	-0.491
	[11.685]	(1.761)	(0.710)	(0.734)	(0.667)
1(children)	0.774	0.033	-0.002	-0.014	-0.012
	[0.418]	(0.053)	(0.027)	(0.028)	(0.027)
Digit span recall	3.020	-0.156	-0.006	-0.010	-0.002
	[1.391]	(0.183)	(0.089)	(0.091)	(0.088)
Years of education	12.111	-0.468	0.161	-0.025	-0.186
	[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	-0.031
	[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	0.183	0.193	0.271

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: 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	-25.90	-14.08	-14.08	-80.66	-45.76
	(21.77)	(21.77)	(0.97)	(0.97)	(18.31)	(14.00)
High treatment	6.77	6.77	-3.98	-3.98	-28.11	1.30
	(24.91)	(24.91)	(0.88)	(0.88)	(19.55)	(15.92)
Control mean	330.97	330.97	78.68	78.68	221.79	228.62
Pr(High treatment = low treatment)	0.08	0.08	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 3: Value of a statistical life: Primary estimates

Panel A: Full sample

	(1) Interacted	(2) Treatment only	(3) Interacted	(4) Treatment only	(5) Know victim	(6) Treatment + Know victim
VSL	215.23	347.26	223.90	307.88	559.99	510.63
	(98.38)	(259.49)	(97.88)	(262.00)	(225.82)	(189.43)
Cragg-Donald F-stat	40.12	10.99	40.75	10.63	16.24	6.32
Weak IV Robust Confidence Set	[24.10, 420.73]	[-160.17, 1,066.93]	[34.22, 428.34]	[-215.85, 1,028.44]	[117.39, 1,002.59]	[188.10, 1,238.20]
Inversion test	CLR	CLR	CLR	CLR	Wald	CLR
Observations	1,425	1,425	1,425	1,425	1,427	1,427
Controls	BL Risk	BL Risk	LASSO	LASSO	None	None
Enumerator FE	Yes	Yes	Yes	Yes	No	No

Panel B: Treated respondents only

	(1) Interacted	(2) Treatment only	(3) Interacted	(4) Treatment only	(5) Know victim	(6) Treatment + Know victim
VSL	396.03	774.94	394.52	823.28	824.37	769.37
	(253.28)	(478.06)	(257.04)	(513.43)	(405.92)	(343.19)
Cragg-Donald F-stat	16.84	11.24	16.07	9.90	8.17	5.68
Weak IV Robust Confidence Set	[-80.04, 1,021.81]	[-39.13, 2,403.92]	[-88.12, 1,038.34]	[-36.41, 2,726.94]	[28.78, 1,619.95]	[239.42, 2,087.47]
Inversion test	CLR	AR	CLR	AR	Wald	CLR
Observations	982	982	982	982	983	983
Controls	BL Risk	BL Risk	LASSO	LASSO	None	None
Enumerator FE	Yes	Yes	Yes	Yes	No	No

Panel C: Interacted estimates leaving 1 arm out

	(1) High treatment	(2) Low treatment	(3) Control	(4) High treatment	(5) Low treatment	(6) Control
VSL	211.18	138.12	396.03	213.28	178.31	394.52
	(88.71)	(190.17)	(253.28)	(87.39)	(174.54)	(257.04)
Cragg-Donald F-stat	79.07	19.80	16.84	82.35	23.50	16.07
Weak IV Robust Confidence Set	[38.16, 391.40]	[-39.13, 2,403.92]	[-80.04, 1,021.81]	[43.23, 390.86]	[-170.17, 559.50]	[-88.12, 1,038.34]
Inversion test	CLR	CLR	CLR	CLR	CLR	CLR
Observations	963	905	982	963	905	982
Controls	BL Risk	BL Risk	BL Risk	LASSO	LASSO	LASSO
Enumerator FE	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parenthesis.

Panel A reports regressions using observations in all experimental arms. Panel B excludes control observations. 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. Panel C reports "interacted" estimates leaving one experimental arm out, listed in the column title. All models control for baseline risk beliefs. Column (5) of panels A and B report estimates that instrument for beliefs using an indicator equal to 1 if the respondent knows a motorcycle accident victim, and this value plus treatment assignment in column (6). Heteroskedastic-robust standard errors are used for these estimates. 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). Just-identified models with heteroskedastic errors in columns (5) - (6) report confidence sets constructed via Wald test inversion. Excludes 35 observations where surveyors reported motorcycle taxi drivers pretending to be passengers.

Table 4: Heterogeneity in the value of a statistical life

	(1) 1(Age > median)	(2) asinh(Wage)	(3) 1(Wage > median)	(4) 1(health > median)	(5) 1(children)	(6) 1(digit recall score > median)	(7) 1(education > median)	(8) 1(female)
VSL	286.64	340.44	8.11	42.85	250.85	60.51	214.97	223.03
	(148.10)	(155.09)	(186.16)	(169.85)	(123.04)	(197.61)	(181.00)	(119.15)
VSL x Interaction	4.39	386.85	501.51	497.87	-97.05	297.54	70.90	260.45
	(268.81)	(127.27)	(273.24)	(254.27)	(247.69)	(242.17)	(232.93)	(238.62)
	[0.987]	[0.002]	[0.066]	[0.050]	[0.695]	[0.219]	[0.761]	[0.275]
VSL elasticity		1.14						
		(0.49)						
Cragg-Donald F-stat	17.72	12.37	18.78	17.01	21.70	12.10	18.78	16.78
Sanderson-Windmeijer first stage F-stats								
Base	36.18	19.58	33.87	48.22	31.56	16.14	25.07	27.87
Interaction	25.10	22.48	26.66	28.34	42.31	22.06	30.73	36.53
Observations	1,423	1,408	1,408	1,425	1,425	1,425	1,417	1,425
Controls	LASSO	LASSO	LASSO	LASSO	LASSO	LASSO	LASSO	LASSO

Standard errors in parenthesis. P-values in brackets.

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. The low treatment and control arms are pooled since posterior beliefs are similar among these groups, and once splitting on some dimensions of heterogeneity, the sample of respondents within certain groups otherwise becomes small since randomization was not stratified. In columns 1, 3-4, and 6-7, the demographic variable is converted to an indicator equal to 1 if the response was above the median. Unemployed individuals have wages coded to 0 in columns 2 and 3 since unemployment is typically involuntary in this sample. The estimates exclude 35 observations for which surveyors reported motorcycle taxi drivers pretending to be passengers.

Table 5: Correlates with beliefs

	(1) 10,000 × Risk	(2) 10,000 × Risk	(3) 10,000 × Risk	(4) 10,000 × Risk	(5) 10,000 × Risk
Panel A: No covariates					
Trips/week	-0.13				
	(3.71)				
Trip length		-1.01			
		(1.80)			
Previous accident			136.02		
			(47.02)		
Contact in accident				266.14	
				(37.21)	
Raining					301.83
					(80.32)
Sample average	333.22	333.22	333.22	333.22	333.22
Observations	1,427	1,427	1,427	1,427	1,427
Panel B: Taxi terminal FEs and controls					
Trips/week	-0.27				
	(3.98)				
Trip length		0.77			
		(1.65)			
Previous accident			90.05		
			(49.74)		
Contact in accident				204.19	
				(58.20)	
Raining					302.55
					(111.69)
Sample average	333.22	333.22	333.22	333.22	333.22
Observations	1,427	1,427	1,427	1,427	1,427

Robust standard errors in parenthesis.

Table 5 reports the correlation of demographic variables with prior beliefs about one's risk of dying in a motorcycle accident over a 5 year span. Estimates in panel a do not include any controls or fixed effects. Estimates in panel b 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. Rainfall is coded to 1 if Visual Crossing reported positive precipitation from a weather station in Nairobi on the day the survey was conducted.

Table 6: Value of a statistical life: Estimates using alternative methods

Panel A: 42% helmet effectiveness treated as truth

	Rational expectations					Subjective beliefs
	(1) OLS	(2) LASSO	(3) LASSO	(4) LASSO	(5) $\frac{WTP_i}{\Delta r_i^*}$	(6) Beliefs
VSL	15,244.30	2,744.41	-6,167.09	1,490.08	381,326.62	110.96
	(10,900.69)	(10,630.65)	(10,840.79)	(10,747.15)	(13,859.27)	(53.74)
95% CI/Confidence set	[-6,121.06, 36,609.66]	[-18091.67, 23,580.49]	[-27415.04, 15,080.86]	[-19574.34, 22,554.50]	[354,162.44, 408,490.79]	[5.63, 216.29]
Cragg-Donald F-stat						19.73
Observations	1,425	1,425	1,425	1,425	1,425	1,425
Controls	None	DPLASSO	DPLASSO	DPLASSO	None	None
Enumerator FE	No	No	Yes	No	No	Yes
Taxi Terminal FE	No	No	No	Yes	No	No

Panel B: 70% helmet effectiveness treated as truth

	Rational expectations					Subjective beliefs
	(1) OLS	(2) LASSO	(3) LASSO	(4) LASSO	(5) $\frac{WTP_i}{\Delta r_i^*}$	(6) Beliefs
VSL	9,146.58	1,646.65	-3,700.25	894.05	228,795.97	110.96
	(6,540.42)	(6,378.39)	(6,504.48)	(6,448.29)	(8,315.56)	(53.74)
95% CI/Confidence set	[-3,672.63, 21,965.79]	[-10855.00, 14,148.30]	[-16449.03, 9,048.52]	[-11744.61, 13,532.70]	[212,497.47, 245,094.48]	[5.63, 216.29]
Cragg-Donald F-stat						19.73
Observations	1,425	1,425	1,425	1,425	1,425	1,425
Controls	None	DPLASSO	DPLASSO	DPLASSO	None	LASSO
Enumerator FE	No	No	Yes	No	No	Yes
Taxi Terminal FE	No	No	No	Yes	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) reports estimates using subjective beliefs about helmet efficacy estimated using a separate survey measure as an instrument to minimize measurement error without ruling out endogeneity. All models exclude 35 observations for which surveyors reported motorcycle taxi drivers pretending to be passengers.

Table 7: 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 posterior 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 along with weak IV robust confidence sets. 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

$$\begin{aligned} v_i &= \zeta_h + VSL\Delta r_i + X_i'\gamma_0 + \gamma_1 r_{0,i} + \epsilon_i \\ \Delta r_i &= Z_i'\pi + X_i'\pi_c + \pi_r r_{0,i} + \nu_i \end{aligned}$$

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

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

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

Hence, by the Law of Total Variance,

$$Avar(\sqrt{N}(\widehat{VSL} - VSL)) = E[V(r_{0,i})] + Var(\mathbb{E}[\sqrt{N}(\widehat{VSL} - VSL)|r_{0,i}])$$
$$= E[V(r_{0,i})] = \mathbb{E}[\epsilon_i]^2\,\mathbb{E}[(\Delta\hat{r}_i)^2]^{-1}$$

So errors are homoskedastic under homogeneous VSL. Although there is some evidence of heterogeneous VSL in this sample, estimates suggest it is small relative to unobserved determinants of utility. The standard deviation of willingness to pay is about 41, so even if the average deviation of VSL were as large as the mean, the contribution to ϵ_i would be just $.022 \cdot 223$ or 10% of the standard deviation of WTP on average, while the difference in perceived risk reduction from a helmet is well under $.022$ across treatment arms. Hence, the data suggest that standard errors are approximately homoskedastic because the contribution of demand for safety is second order compared to demand for other characteristics of helmets, consistent with the large demand intercept and small coefficient on mortality risk reduction.

Leveraging this approximation allows for efficient VSL estimation using two-stage least squares and lends itself to a well-established literature on weak instrument robust inference under homoskedasticity. 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).

For robustness, I also report estimates constructed using continuously updating GMM (CUE) with heterodekastic robust standard errors in Appendix Table A6. I report estimates using efficient GMM since 2SLS is less efficient under heteroskedasticity and to show that results are similar under different weighting of the instruments. Weak instrument robust confidence sets are reported based on inversion of a CLR test statistic. Standard errors increase, but the primary conclusions are unchanged. The upper bound of confidence sets is about \$700 with the interacted IV and \$1,000 with the treatment only instrument with or without covariates.

C Local average VSLs identified by interacted and non-interacted instruments

This section derives the weighted average VSLs identified by the treatment only and interacted vectors of instruments. I begin with the case where r_{i0} and Δr_i are measured without systematic error. Since agents do not update beliefs about the unhelmeted risk of motorcycles when presented with information, their posterior beliefs about the likelihood that a helmet will save their life is approximately given by

$$\Delta r_i = H_i(T_i) r_{i0} + \nu_i$$

where r_{i0} is their prior about the mortality risk of motorcycles and H_i is the posterior belief about the efficacy of a helmet. Letting L_i indicate assignment to the low treatment arm and H_i to the high treatment arm, the relationship is therefore approximated by

$$\Delta r_i = \pi_c r_{i0} + \pi_L L_i \cdot r_{i0} + \pi_H H_i \cdot r_{i0} + \nu_i$$

One may also estimate a treatment only first stage

$$\Delta r_i = \pi_1 L_i + \pi_2 H_i + \nu_i'$$

And an individual's helmet valuation can be written as

$$v_i = \alpha + VSL_i \Delta r_i + \epsilon_i$$

I will apply the Frisch-Waugh-Lovell theorem to partial out the constant in the treatment only first stage and r_{i0} from the interacted version. I will denote by $\tilde{x}_i$ the partialled out variables.

Beginning with the treatment only IV, substitution of the known form for Δr_i into the structural equation for v_i yields

$$\begin{aligned}
V\hat{S}L^{TO}_{2SLS} &= \frac{Cov(v_i, \pi_1 \tilde{L}_i + \pi_2 \tilde{H}_i)}{Cov(\Delta r_i, \pi_1 \tilde{L}_i + \pi_2 \tilde{H}_i)} \\
&= \frac{Cov(VSL_i \pi_L L_i \cdot r_{i0} + VSL_i \pi_H H_i \cdot r_{i0}, \pi_1 \tilde{L}_i + \pi_2 \tilde{H}_i)}{\pi_L \pi_1 V(L_i)\, \mathbb{E}[r_{i0}] + \pi_H \pi_2 V(H_i)\, \mathbb{E}[r_{i0}]} \\
&= \mathbb{E}\left[VSL_i \frac{r_{i0}\left(\pi_1 \pi_L V(L_i) + \pi_2 \pi_H V(H_i)\right)}{\mathbb{E}[r_{i0}]\left(\pi_1 \pi_L V(L_i) + \pi_2 \pi_H V(H_i)\right)} \right] \\
&= \mathbb{E}\left[VSL_i \frac{r_{i0}}{\mathbb{E}[r_{i0}]} \right]
\end{aligned}$$

Showing the estimate weights the VSL of individuals linearly in their prior, r_{i0}. In the interacted case, observe that first stage coefficients on L_i and H_i will be zero in expectation, therefore

$$\begin{aligned}
V\hat{S}L^{INT}_{2SLS} &= \frac{Cov(v_i, \pi_L \widetilde{r_{i0} \cdot L_i} + \pi_H \widetilde{r_{i0} \cdot H_i})}{Cov(\Delta r_i, \pi_L \widetilde{r_{i0} \cdot L_i} + \pi_H \widetilde{r_{i0} \cdot H_i})} \\
&= \frac{Cov(VSL_i \pi_L L_i \cdot r_{i0} + VSL_i \pi_H H_i \cdot r_{i0}, \pi_L \widetilde{r_{i0} \cdot L_i} + \pi_H \widetilde{r_{i0} \cdot H_i})}{\pi_L^2 V(L_i)\, \mathbb{E}[r_{i0}^2] + \pi_H^2 V(H_i)\, \mathbb{E}[r_{i0}^2]} \\
&= \mathbb{E}\left[VSL_i \frac{r_{i0}^2\left(\pi_L^2 V(L_i) + \pi_H^2 V(H_i)\right)}{\mathbb{E}[r_{i0}^2]\left(\pi_L^2 V(L_i) + \pi_H^2 V(H_i)\right)} \right] \\
&= \mathbb{E}\left[VSL_i \frac{r_{i0}^2}{\mathbb{E}[r_{i0}^2]} \right]
\end{aligned}$$

Therefore the interacted instrument weights observations proportionally to priors squared, r_{i0}^2.

Experimenter demand effects:

Next consider a model of experimenter demand effects where agents report $\Delta r_i^* = \pi_c r_{i0} + \pi_L \zeta_L L_i \cdot r_{i0} + \pi_H \zeta_H H_i \cdot r_{i0} + \nu_i \neq \Delta r_i$. This covers a case where agents over or under-report true changes in beliefs in response to the intervention. Valuations are still a product of true beliefs, Δr_i. In this case,

$$\begin{aligned}
V\hat{S}L^{TO}_{2SLS} &= \mathbb{E}\left[VSL_i \frac{r_{i0}}{\mathbb{E}[r_{i0}]} \right] \cdot \frac{V(L_i)\pi_1 \pi_L \zeta_L + V(H_i)\pi_2 \pi_H \zeta_H}{V(L_i)\pi_1 \pi_L \zeta_L^2 + V(H_i)\pi_2 \pi_H \zeta_H^2} \\
V\hat{S}L^{INT}_{2SLS} &= \mathbb{E}\left[VSL_i \frac{r_{i0}^2}{\mathbb{E}[r_{i0}^2]} \right] \cdot \frac{V(L_i)\pi_L^2 \zeta_L + V(H_i)\pi_H^2 \zeta_H}{V(L_i)\pi_L^2 \zeta_L^2 + V(H_i)\pi_H^2 \zeta_H^2}
\end{aligned}$$

Hence, $V\hat{S}L^{TO}_{2SLS} \neq V\hat{S}L^{INT}_{2SLS}$ unless $\zeta_L = \zeta_H$ since I reject $\pi_L = \pi_H$ in the data.

Misreported mortality risk:

Suppose that agents report $r_{i0}\zeta_i$ for some $\zeta_i > 0$ due to systematic measurement error but the true beliefs guiding valuations are r_{i0}. This captures challenges reporting small probabilities. One can show that

$$\hat{VSL}_{2SLS}^{TO} = \mathbb{E}\left[VSL_i \frac{r_{i0}}{\mathbb{E}[r_{i0}\zeta_i]}\right]$$
$$\hat{VSL}_{2SLS}^{INT} = \mathbb{E}\left[VSL_i \frac{r_{i0}^2\zeta_i}{\mathbb{E}[r_{i0}^2\zeta_i^2]}\right]$$

Therefore one typically has $\hat{VSL}_{2SLS}^{TO} \neq \hat{VSL}_{2SLS}^{INT}$ unless ζ_i and r_{i0} are independent, for instance if $\zeta_i = c$. Observe that with "S-shaped" weighting as documented in prospect theory, $\mathbb{E}[\zeta_i r_{i0}] > 0$. Similarly, if agents round small probabilities up or down, $\mathbb{E}[\zeta_i r_{i0}] \neq 0$.

D Appendix Figures

A1: Survey locations

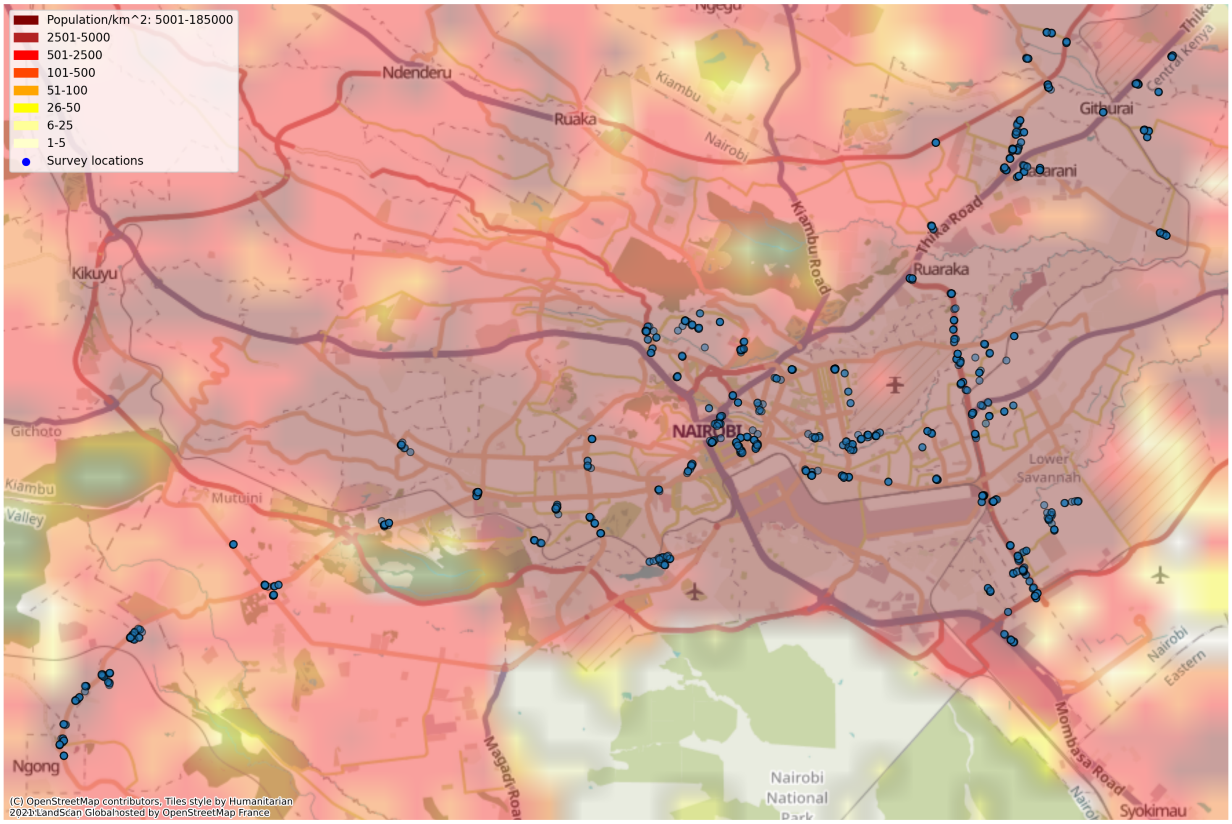
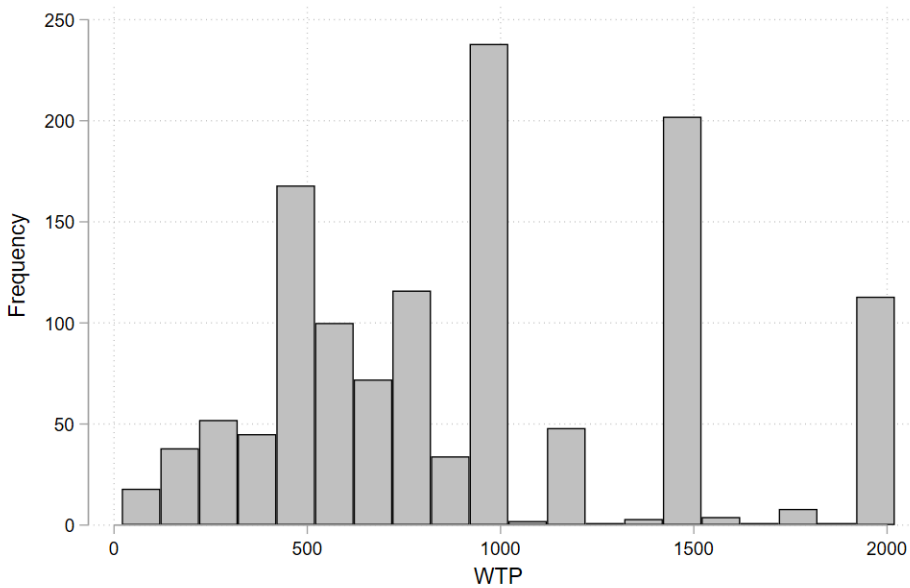


Figure A1 plots survey locations over a map of Nairobi. Color denotes population, where no color denotes no residents. The 2019 census estimated a population of about 4.4 million within the city. Map data is from Open Street Map. Population data is from the LandScan Global 2021 data set produced by Oak Ridge National Laboratory.

A2: Distribution of helmet bids (Kenyan shillings)

(a) A. Histogram of bids, restricted axis



(b) B. Histogram of bids, full distribution

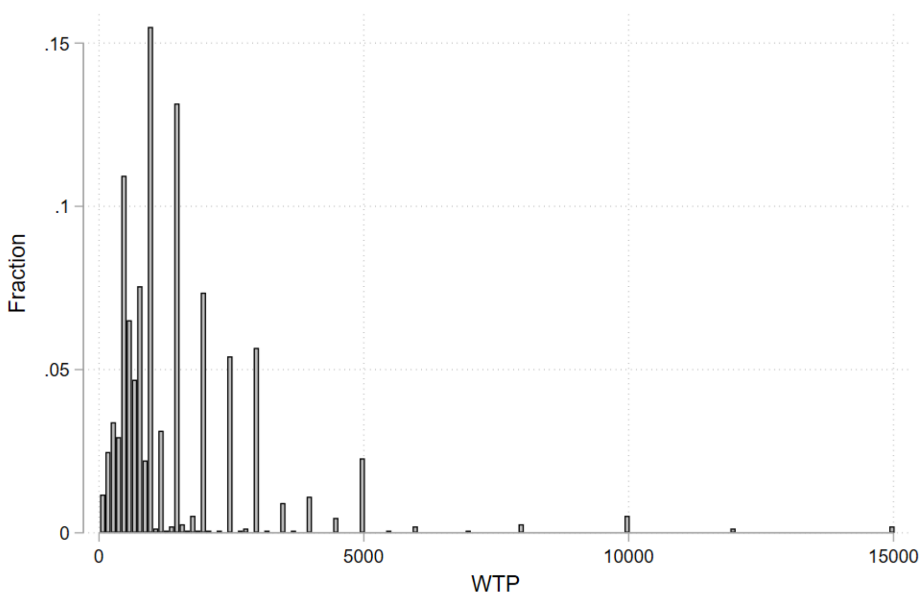


Figure A2 plots the distribution of willingness to pay for helmets in Kenyan shillings (Ksh). Panel A reports a histogram of bids, excluding outliers above Ksh 2,000 for clarity. Panel B reports the same histogram across the full distribution of bids. Figures present Ksh rather than USD to illustrate that respondents are more likely to select round numbers.

A3: VSL estimates across studies, not truncated

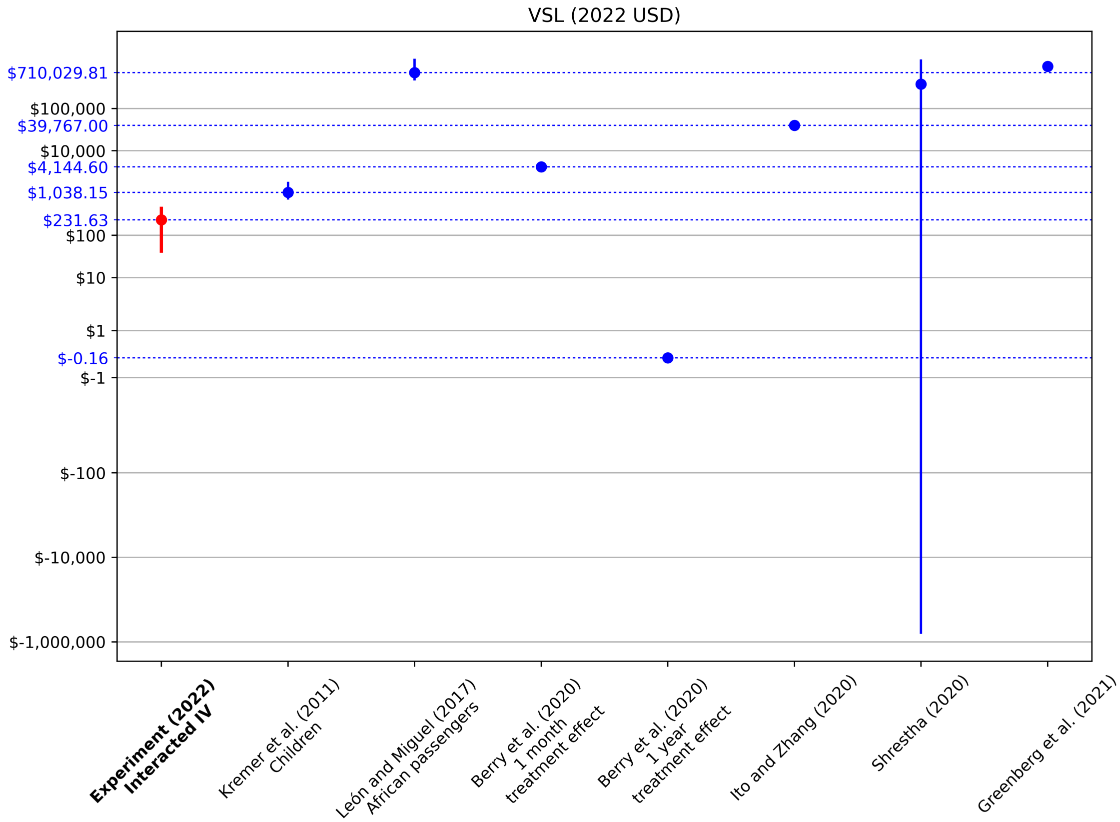


Figure A3 plots revealed-preference VSL estimates and (where available) 95% confidence intervals from this paper, Kremer et al. (2011), León and Miguel (2017), Berry et al. (2020), Ito and Zhang (2020), Shrestha (2020), and Greenberg et al. (2021). Greenberg et al. (2021) is included for comparison to a high-income setting. The other estimates are from low and middle income economies. All estimates are presented in 2022 USD calculated by inflating based on the paper's publication year using the CPI inflation calculator.

A4: Effect of study VSL on published benefit-cost ratios

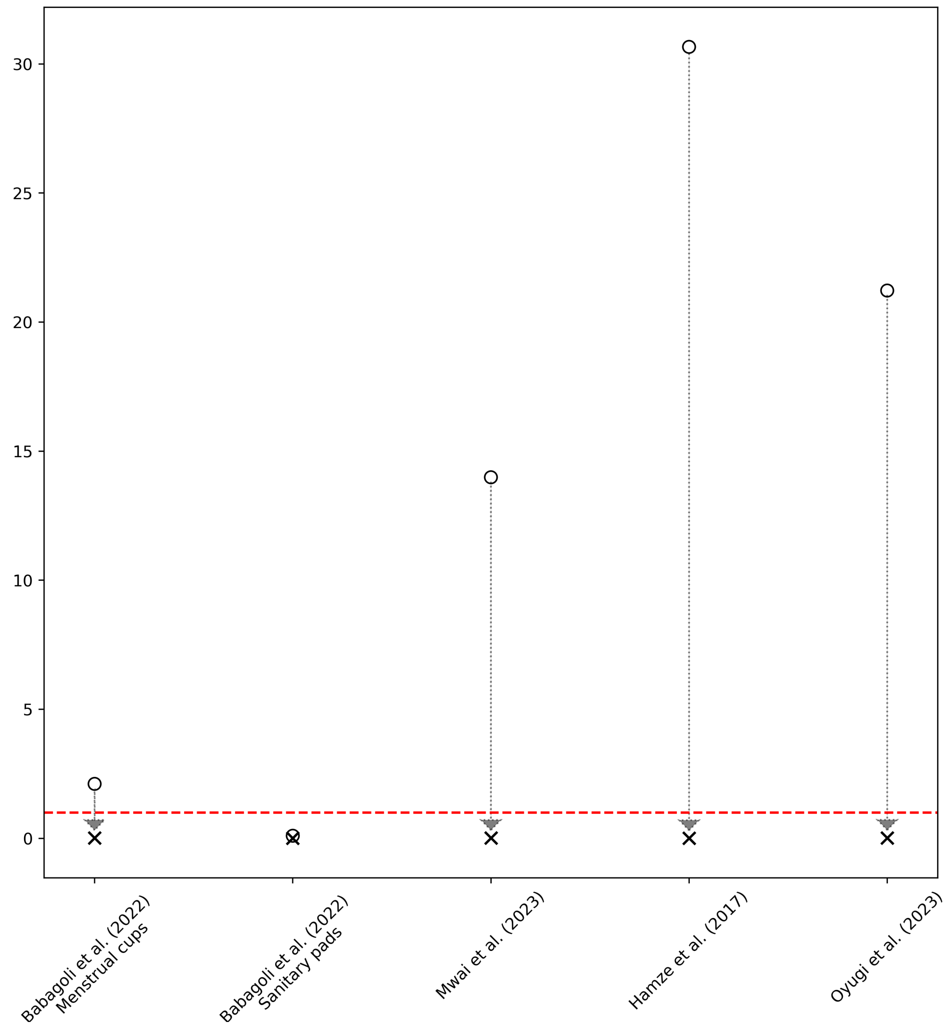


Figure A4 examines how the benefit-cost ratios (BCRs) of published benefit-cost analyses of Kenyan programs change when the study's original VSL or VSLY is replaced with the preferred estimates from this study. Hollow circles denote the original BCR estimate, and an x denotes the revised estimate. The 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.

E Appendix Tables

A1: 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	0.196	-0.379	-0.593*
	[5.242]	(0.725)	(0.348)	(0.358)	(0.343)
Average trip length (minutes)	19.593	0.334	-0.704	-1.301*	-0.583
	[12.567]	(1.538)	(0.725)	(0.745)	(0.672)
$\mathbb{E}$[Deaths/10,000 passengers, 1 year]	371.584	NA	-46.930	-0.326	46.539
	[1,186.706]	NA	(76.351)	(78.569)	(76.478)
$\mathbb{E}$[Deaths/10,000 passengers, 5 years]	977.118	NA	-154.821	266.926	420.383**
	[2,479.942]	NA	(192.351)	(197.928)	(205.063)
Confidence in beliefs	3.461	NA	0.034	0.011	-0.022
	[0.699]	NA	(0.042)	(0.044)	(0.041)
10000*Pr(Fatal accident, 5 years)	354.133	NA	-19.377	-40.568	-20.464
	[973.710]	NA	(56.011)	(57.630)	(52.741)
Previous accident	0.488	NA	-0.011	-0.030	-0.019
	[0.500]	NA	(0.032)	(0.033)	(0.032)
Know accident victim	0.947	NA	-0.022	-0.036**	-0.014
	[0.224]	NA	(0.017)	(0.017)	(0.017)
Use motorcycle taxi: Commuting	0.781	0.027	-0.032	0.008	0.038
	[0.414]	(0.054)	(0.027)	(0.028)	(0.027)
Shopping	0.420	-0.031	-0.033	-0.018	0.016
	[0.494]	(0.064)	(0.031)	(0.032)	(0.031)
Leisure	0.261	0.057	0.074**	0.036	-0.037
	[0.440]	(0.058)	(0.029)	(0.030)	(0.029)
Deliveries	0.095	-0.012	0.015	-0.000	-0.015
	[0.294]	(0.014)	(0.018)	(0.018)	(0.018)
Emergency/hospital transportation	0.095	-0.008	-0.009	-0.007	0.003
	[0.294]	(0.014)	(0.017)	(0.017)	(0.017)
Reason for use: Speed	0.816	0.114**	0.019	0.003	-0.017
	[0.388]	(0.045)	(0.024)	(0.025)	(0.024)
Convenience	0.717	-0.005	0.005	-0.045	-0.049*
	[0.451]	(0.061)	(0.029)	(0.030)	(0.029)
Only option	0.168	0.007	-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*	-0.010	0.017
	[0.253]	(0.033)	(0.015)	(0.015)	(0.014)
Enjoyment	0.011	-0.012	-0.002	0.008	0.009
	[0.105]	(0.015)	(0.007)	(0.007)	(0.007)
Risk information: Own experiences	0.819	NA	-0.036	-0.042	-0.007
	[0.386]	NA	(0.026)	(0.027)	(0.026)
Friends/family	0.454	NA	0.029	0.038	0.009
	[0.498]	NA	(0.032)	(0.033)	(0.032)
Social media	0.414	NA	-0.049	-0.034	0.016
	[0.493]	NA	(0.031)	(0.032)	(0.031)
Media	0.288	NA	0.014	-0.027	-0.042
	[0.453]	NA	(0.029)	(0.029)	(0.028)
Government	0.135	NA	0.015	0.026	0.011
	[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	0.136	0.221	0.221

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.

A2: 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	0.004	0.003	-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)
Income (PPP, ’000s USD)	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**	0.016
	[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	0.419	0.145	0.660

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: 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, 1 year]	0.004	NA	0.001	0.006	0.005
	[0.066]	NA	(0.005)	(0.005)	(0.006)
$\mathbb{E}$[Deaths/10,000 passengers, 5 years]	0.004	NA	0.001	0.006	0.005
	[0.066]	NA	(0.005)	(0.005)	(0.006)
10000*Pr(Fatal accident, 5 years)	176.991	NA	10.834	56.171	43.751
	[1,320.015]	NA	(89.528)	(92.021)	(90.661)
Confidence in beliefs	0.002	NA	-0.002	0.002	0.004
	[0.047]	NA	(0.003)	(0.003)	(0.003)
Previous accident	0.002	NA	-0.002	-0.002	0.000
	[0.047]	NA	(0.002)	(0.002)	(0.000)
Know accident victim	0.000	NA	0.000	0.000	0.000
	[0.000]	NA	(0.000)	(0.000)	(0.000)
Use motorcycle taxi: Commuting	0.000	0.000	0.000	0.000	0.000
	[0.000]	(0.000)	(0.000)	(0.000)	(0.000)
Shopping	0.000	0.000	0.000	0.000	0.000
	[0.000]	(0.000)	(0.000)	(0.000)	(0.000)
Leisure	0.000	0.000	0.000	0.000	0.000
	[0.000]	(0.000)	(0.000)	(0.000)	(0.000)
Deliveries	0.000	0.000	0.000	0.000	0.000
	[0.000]	(0.000)	(0.000)	(0.000)	(0.000)
Emergency/hospital transportation	0.000	0.000	0.000	0.000	0.000
	[0.000]	(0.000)	(0.000)	(0.000)	(0.000)
Reason for use: Speed	0.000	0.000	0.000	0.000	0.000
	[0.000]	(0.000)	(0.000)	(0.000)	(0.000)
Convenience	0.000	0.000	0.000	0.000	0.000
	[0.000]	(0.000)	(0.000)	(0.000)	(0.000)
Only option	0.000	0.000	0.000	0.000	0.000
	[0.000]	(0.000)	(0.000)	(0.000)	(0.000)
Price	0.000	0.000	0.000	0.000	0.000
	[0.000]	(0.000)	(0.000)	(0.000)	(0.000)
Safety/Avoid dangerous areas	0.000	0.000	0.000	0.000	0.000
	[0.000]	(0.000)	(0.000)	(0.000)	(0.000)
Enjoyment	0.000	0.000	0.000	0.000	0.000
	[0.000]	(0.000)	(0.000)	(0.000)	(0.000)
Risk information: Own experiences	0.000	NA	0.000	0.000	0.000
	[0.000]	NA	(0.000)	(0.000)	(0.000)
Friends/family	0.000	NA	0.000	0.000	0.000
	[0.000]	NA	(0.000)	(0.000)	(0.000)
Social media	0.000	NA	0.000	0.000	0.000
	[0.000]	NA	(0.000)	(0.000)	(0.000)
Media	0.000	NA	0.000	0.000	0.000
	[0.000]	NA	(0.000)	(0.000)	(0.000)
Government	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		.	0.500	0.465	0.534

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.

A4: Effect of information on beliefs: Estimates 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	-20.56	-14.06	-14.06	-76.75	-40.96
	(21.72)	(21.72)	(0.97)	(0.97)	(18.36)	(14.14)
High treatment	7.58	7.58	-4.12	-4.12	-26.99	2.41
	(24.44)	(24.44)	(0.87)	(0.87)	(19.16)	(15.64)
Control mean	339.71	339.71	78.85	78.85	226.92	233.96
Pr(High treatment = low treatment)	0.13	0.13	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.

A5: 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	3.18	5.41	3.39	5.27	1.47	4.50
	(1.52)	(4.62)	(1.52)	(4.64)	(1.11)	(3.43)
Cragg-Donald F-stat	54.55	10.80	55.91	10.92	55.91	10.92
Weak IV Robust Confidence Set	[0.24, 6.32]	[-3.70, 18.55]	[0.45, 6.53]	[-3.89, 18.05]	[-0.70, 3.71]	[-2.16, 13.30]
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	5.55	16.65	4.82	15.23	1.89	7.78
	(4.01)	(10.64)	(3.95)	(10.05)	(2.77)	(6.67)
Cragg-Donald F-stat	23.75	8.78	24.06	9.32	24.06	9.32
Weak IV Robust Confidence Set	[-2.10, 14.90]	[-0.84, 60.48]	[-2.80, 13.89]	[-1.69, 53.71]	[-3.65, 7.94]	[-4.54, 31.07]
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.

A6: Value of a statistical life: Estimates with heteroskedastic robust standard errors

	Raw WTP				Winsorized WTP	
	(1) Interacted	(2) Treatment only	(3) Interacted	(4) Treatment only	(5) Interacted	(6) Treatment only
VSL	249.34	349.35	245.74	308.12	102.70	252.59
	(138.64)	(233.33)	(141.87)	(238.51)	(76.48)	(193.40)
Cragg-Donald F-stat	8.23	11.84	8.16	10.96	8.16	10.96
Weak IV Robust Confidence Set Inversion test	[-19.64, 704.97] CLR	[-103.35, 1,005.30] CLR	[-40.74, 689.51] CLR	[-154.63, 978.63] CLR	[-63.84, 323.75] CLR	[-122.64, 750.35] 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

Robust standard errors in parenthesis and weak instrument robust confidence sets in brackets.

Table A6 reports VSL estimates under heteroskedastic robust standard errors using continuous updating GMM (CUE). Weak instrument robust confidence sets are calculated using conditional likelihood ratio test inversion. 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.

A7: 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	39.86	273.66	59.07	211.15	44.53	232.41
	(98.86)	(278.59)	(99.06)	(282.85)	(70.46)	(204.71)
Cragg-Donald F-stat	41.88	9.99	42.05	9.57	42.05	9.57
Weak IV Robust Confidence Set	[-159.33, 240.49]	[-297.40, 1,054.76]	[-159.33, 240.49]	[-395.81, 984.34]	[-96.19, 187.12]	[-178.22, 770.64]
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	29.28	761.86	-26.67	779.79	26.24	424.19
	(251.24)	(534.02)	(256.07)	(568.27)	(175.23)	(369.34)
Cragg-Donald F-stat	17.18	9.63	16.16	8.46	16.16	8.46
Weak IV Robust Confidence Set	[-516.47, 609.45]	[-158.62, 2,727.95]	[-605.12, 556.92]	[-194.57, 3,080.62]	[-351.33, 418.50]	[-259.50, 1,809.29]
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.

A8: 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	141.57	287.46	154.61	264.60	66.63	199.55
	(64.91)	(187.29)	(65.85)	(195.50)	(48.09)	(144.70)
Cragg-Donald F-stat	45.37	10.37	44.74	9.52	44.75	9.52
Weak IV Robust Confidence Set	[13.66, 274.52]	[-82.08, 803.62]	[25.46, 290.05]	[-130.39, 811.47]	[-28.66, 164.14]	[-86.79, 578.53]
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	255.77	332.13	277.79	298.61	144.55	240.87
	(89.62)	(245.11)	(89.07)	(246.66)	(64.47)	(182.05)
Cragg-Donald F-stat	41.49	10.38	42.24	10.06	42.24	10.06
Weak IV Robust Confidence Set	[83.72, 445.25]	[-151.15, 1,032.40]	[107.83, 466.78]	[-198.32, 988.80]	[19.06, 277.94]	[-118.93, 712.36]
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 C: Weighted by 1/rides per week

	Raw WTP				Winsorized WTP	
	(1) Interacted	(2) Treatment only	(3) Interacted	(4) Treatment only	(5) Interacted	(6) Treatment only
VSL	272.75	96.56	286.78	86.40	145.99	110.97
	(100.91)	(311.00)	(98.70)	(304.39)	(72.20)	(225.65)
Cragg-Donald F-stat	43.20	7.97	45.41	8.26	45.41	8.26
Weak IV Robust Confidence Set	[77.43, 482.10]	[-637.11, 865.77]	[96.84, 492.58]	[-632.16, 837.48]	[4.88, 293.38]	[-388.60, 671.41]
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 accounts 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 uses the respondent's stated belief about the lifespan of the helmet, rather than the manufacturers suggestion. Panel C weights each observation by the inverse of motorcycle trips in a typical week to account for selection into ridership. 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.

A9: 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	534.88	671.97	601.70	653.89	282.14	436.77
	(252.81)	(416.68)	(280.05)	(452.65)	(195.98)	(326.19)
Cragg-Donald F-stat	9.54	7.38	8.03	6.23	8.03	6.23
Weak IV Robust Confidence Set	[64.21, 1,197.10]	[-90.81, 2,026.93]	[64.21, 1,197.10]	[-187.78, 2,297.45]	[-108.59, 772.92]	[-194.93, 1,492.94]
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	676.59	969.44	690.84	1,022.45	323.45	535.40
	(401.69)	(595.22)	(414.44)	(633.37)	(281.93)	(420.67)
Cragg-Donald F-stat	10.67	10.97	9.95	9.83	9.95	9.83
Weak IV Robust Confidence Set	[-66.05, 1,785.13]	[-49.48, 3,012.90]	[-72.23, 1,869.17]	[-45.73, 3,365.04]	[-237.52, 1,067.29]	[-239.47, 1,959.91]
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.