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

May 14, 2025

Abstract

Economists often study non-market goods such as health and air quality. This paper introduces a new method to estimate demand for such amenities and applies it to measure the value of a statistical life (VSL) in Kenya. My approach is to update beliefs about the life-saving efficacy of a product (a motorcycle helmet) and elicit product choice. This generates instruments allowing one to use subjective beliefs to estimate demand, rather than assuming rational expectations. This method does not require beliefs to be reported error-free but does require classical mismeasurement. I validate this assumption using features of the experimental design. The estimated VSL is \$224, near the left tail of Kenyan estimates. Standard methods for estimating VSL produce skewed results, driven by severe violations of rational expectations. These findings help explain low observed demand for many health products and suggest that directing more development aid towards consumption may increase welfare.

Keywords: Non-market amenities, value of a statistical life, rational expectations

*I thank Edward Miguel and Supreet Kaur for excellent feedback and advising. This project benefited from suggestions by Paul Niehaus, Patrick Kline, Christopher Snyder, Kurt Lavetti, Dennis Egger, and Kelsey Jack. I am grateful to the UC Berkeley development and industrial organization communities and to seminar and conference participants for valuable feedback. 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, 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, products often change exposure to or consumption of such goods (e.g. air 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. It also does not require that agents fully update beliefs in response to information, 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. 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 challenging to estimate. Mortality risk is a feature of economic decisions including job choice, healthcare, environmental regulation and transportation. Policymakers 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 paternalism. 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 models where the level of mortality risk present in data is used as a proxy for beliefs. This approach is biased if agents have beliefs about mortality risk that differ from statistical estimates (i.e. they do not have rational expectations). Moreover, estimates may be prone to omitted variable bias if agents select into dangerous behavior (Ashenfelter and Greenstone, 2004). VSL may also be heterogeneous, and many studies estimate it over selected populations where there is quasi-random variation in 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).[2]

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 the choice between a helmet or cash. The approach does not rely on rational expectations, and the random assignment of information ensures that selection into risk does not bias estimates. Respondents were assigned to an information control or one of two treatments. One treatment arm was presented with 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 risk information, but a 70% helmet effectiveness estimate (Ouellet and Kasantikul, 2006), which differs because experts are uncertain about how well helmets work. I then elicited respondents' beliefs about the likelihood that a helmet would save their life and estimated helmet demand using a Becker et al. (BDM,

[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.

[2]Stated preference VSL elicitation does not assume rational expectations, but agents have no incentivize to answer truthfully and estimates may be prone to desirability bias, so I focus on revealed preference methods.

1964) mechanism, inverted to offer a free helmet or cash so that demand is not limited by cash on hand. I estimate VSL by fitting a 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 helmet use is low, enabling VSL estimation in a minimally selected population. Traffic accidents are a leading cause of death in Kenya and motorcycle taxi use is widespread. Yet only 3% of passengers wore helmets near the time of the study (Bachani et al., 2017). Since fatality risk rises with trip volume, the setting also facilitates comparison with traditional VSL estimation approaches.

In the first main finding of the study, I estimate that Kenyans have low demand for marginal reductions in mortality risk, with a VSL of PPP USD $224 (5% of annual income).[3] Receipt of the 42% helmet effectiveness estimate lowers average beliefs about the likelihood that a helmet will save one's life (a product of accident risk and helmet efficacy) from 2.22% to 1.41%, reducing demand for a helmet by $3. This approach recover informative bounds on VSL, rejecting values below $34 and above $429 with 95% confidence.

The low VSL estimates are consistent with economic theory. VSL is inversely proportional to the marginal utility of consumption and should rise with income as the opportunity cost of funds falls. The gap between this estimate and a $700,000 VSL (León and Miguel, 2017), among a sample with incomes near $75,000, can be fully explained by macroeconomic estimates of utility curvature (Havránek, 2015).[4] I also estimate heterogeneity that aligns with theory: VSL is $500 higher among wealthier individuals in the sample, with an estimated income elasticity of 1.14, and higher among healthier respondents.

Is it plausible that the VSL in this sample is below annual income? This finding aligns with a literature documenting low demand for preventative health in LMICs. For instance, Cohen and Dupas (2010) find that only 40% of pregnant women purchase insecticide treated malaria nets at a price of $0.6, implying a cost per life saved under $200 (Pryce et al., 2018). Similarly, Banerjee et al. (2010) show that $4 in in-kind incentives sub-

[3] I use the World Bank's 2021 PPP conversion factor to convert to USD (43.8).

[4] I selected this paper because it considers a population in Africa, making it comparable on dimensions other than income. VSL estimates from the United States are higher still.

stantially increase childhood vaccination. Low demand could, of course, reflect decision-making errors or binding constraints. But adoption has often remained low even when capital, information, and behavioral explanations are targeted (Dupas and Miguel, 2017). This study suggests that low adoption may simply reflect low willingness to pay for marginal reductions in mortality risk in contexts where the opportunity cost of funds is high.

The second component of this study validates that beliefs are measured without non-classical error. Two leading threats are experimenter demand effects and systematically misreported beliefs. Since the concern is a gap between reported and decision-relevant subjective beliefs, not accuracy relative to empirical truth, tools like bunching tests are uninformative. But two design features help test for bias. First, the two treatments provide signals of varying intensity. If beliefs have classical error, VSL estimates should be unchanged if one arm is excluded. Otherwise one would expect values to differ since posteriors are centered at different points across treatments. I show that using any combination of arms to instrument for beliefs yields similar estimates. Moreover, helmet demand falls in response to the low signal, providing revealed preference evidence of belief updating.

The other design feature that allows for tests of misreported beliefs draws on the logic of local average treatment effects. Instrumenting for posteriors using treatment assignment weights observations linearly in priors, but one may construct instruments that weight observations proportionally to priors squared, yielding similar results. The stability of estimates suggests that estimates are not substantially affected by misreporting correlated with true beliefs, which I show includes most cases where agents systematically under or over-report probabilities (e.g. rounding or difficulty articulating small probabilities).

Additional checks further support the validity of the results. First, I proxy for perceived mortality risk using respondents' estimates of deaths per 10,000 motorcycle riders over a helmet's lifespan. This approximates the elicited probability used in the main estimates but avoids low probability reports. VSL estimates are similar, suggesting challenges articulating rare events do not drive results. VSL also does not vary significantly with education or digit span recall. Second, helmet demand is similar between the control and a pure control not asked questions about risk, and instrumenting for beliefs using an observational shifter (personally knowing an accident victim) yields similar VSL estimates, inconsistent with Hawthorne effects. Finally, the BDM mechanism was implemented in a high-stakes, realistic setting where evidence supports its reliability (Berry et al., 2020).

The third part of the paper tests if rational expectations, a core assumption in most VSL estimation methods, holds in this context. The data suggest systematic departures from rational expectations because agents learn from their experiences and those of their social network, not from representative data. Those that know someone involved in a motorcycle accident perceive higher 5-year fatal accident risk, but those that ride motorcycles frequently do not. Respondents also report twice the risk if it rained the day of the survey, raising the salience of risk. Such deviations from rational expectations are likely to bias VSL estimates from typical approaches. Consistent with this prediction, two common methods produce results outside of confidence sets obtained from the new approach.

These results have important policy implications. The VSL estimates are similar to the lower range of estimates from LMICs (Kremer et al., 2011), but they are much smaller than the values used in 5 recent academic benefit-cost analyses. In 4 out of 5 cases, they cause benefit-cost ratios to fall below 1. 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, possibly leading to a misallocation of foreign aid.

The findings should not be interpreted as a call to eliminate health funding for LMICs. The low VSL highlights that the parameter only reflects preferences for marginal mortality risk reduction versus consumption. For many health policies (e.g. WHO funding allocation), decisions are over whom to distribute resources to, and the consumption of recipients is unaffected. In these cases, consumers do not face health-consumption tradeoffs, so applying VSL would undervalue welfare gains of poor beneficiaries. VSL also should not be used to value medical treatments (e.g. antiretrovirals) that produce large reductions in mortality risk. The transfers needed to compensate individuals for such risks may lower their marginal utility of consumption, increasing demand when risks are non-marginal. Finally, many health interventions yield additional benefits, such as lower exposure to non-fatal morbidities or positive externalities onto non-recipients, that are not captured by VSL.

This paper is among the first to estimate VSL using subjective beliefs and with experimental methods that are robust to selection into risk.[5] This addresses two of the leading concerns with traditional methods. The results also provide one of the first tests of rational expectations in the VSL literature. This builds on Ashenfelter and Greenstone (2004) and Greenberg et al. (2021) which study endogeneity and heterogeneity in VSL. It also relates to Baylis et al. (2023) which study rational expectations in the setting of air quality. The VSL estimates are also independently important because knowledge of demand for safety in LMICs is limited (Greenstone and Jack, 2015).[6] I produce one of the first urban estimates using a design that ensures results are not determined by cash on hand, which has caused some to question revealed preference estimates in poor settings (Redfern et al., 2019).

This method could be applied to non-market amenities beyond safety. Goods such as time, privacy, reputation, and environmental quality are economically important but difficult to value (Campbell and Brown, 2003; Greenstone and Jack, 2015). Policy decisions often rely on estimates that are not incentivized (e.g. contingent valuation) when there is no variation for revealed preference measures (Mendelsohn and Olmstead, 2009). Methods also tend to assume rational expectations, which contradicts recent evidence (Baylis et al., 2023). The framework I introduce could potentially estimate demand for such amenities using products such as ride sharing services, cyber security software, or air purifiers.

This research also contributes to a literature on eliciting beliefs for use in economic models (e.g. Manski, 2004; Wiswall and Zafar, 2018). I show how one can design information treatments that produce multiple instruments for beliefs which mitigate measurement error and facilitate tests for bias. Practically, the survey instrument was also effective at measuring expectations. These tools are important for applied work since noisy elicitation has prevented the use of subjective beliefs in some cases (León and Miguel, 2017).

[5] Shrestha (2020) attempts to estimate VSL using similar methods, but confidence sets are uninformative.

[6] 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 relatively 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 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 since it is a function of ridership. This facilitates comparisons to typical methods for estimating VSL.

Motorcycle taxi use is widespread and growing in 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.[7] Transportation is dangerous. Data from the National Transport and Safety Authority (NTSA) reports that 1,722 motorcycle users died in 2021, up from 715 in 2017. 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.[8]

Despite the high risks, helmet use among taxi passengers is rare. Bachani et al. (2017) measure passenger helmet use at 3%. At that time, helmeted passengers typically borrowed one from the driver, which became rare after the COVID-19 pandemic due to hygiene concerns, anecdotally reducing helmet use. Low helmet adoption suggests that demand may be low. However, safe helmets have only recently become available at an affordable price in Kenya. The *FIA Foundation* began offering helmets in Kenya in 2021, about a year before the experiment was implemented, and a local producer began manufacturing helmets near the same time. Retail diffusion of both products was limited, and many consumers reported that they were unaware they could purchase a helmet without a motorcycle.

Do consumers with low safety demand select into riding motorcycles? This setting was chosen because most Kenyans use motorcycle taxis, so selection is limited at the extensive

[7] Fred Matiang'i, "The urgency of bodaboda reforms", *Nation.Africa*, 2022.
[8] "New initiative to tackle road crash deaths in Kenya," *World Health Organization*.

margin. They are used similarly to traditional taxis or ride-sharing vehicles in high-income countries. It is common to take one on trips where public transit is unavailable or when running late, but less common for daily commutes, which are more publicly taken on buses in Nairobi. 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 ridership.[9] Selection may be limited because alternative modes of public transportation are dangerous, crowded and uncomfortable.

Unlike the extensive margin, ridership volume may be more selected. 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 test for selection. Section 5.1 presents details and shows that re-weighting is inconsequential, suggesting that substantial selection on safety preferences is unlikely driving results. The VSL of motorcycle taxi users is also policy relevant even if the population is not fully 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. The stands were selected for broad geographic and demographic coverage. However, areas with high crime rates were excluded for safety reasons (for both survey enumerators and respondents). 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 they could choose a free helmet or a cash gift if they

[9]There were 22 million trips per day in 2022, and a population of 53 million. 40% of Kenyans are farmers and at least 10% are 5 or younger. Respondents report taking 7.5 trips/week on average. This suggests 80% of urban Kenyans adults ride motorcycle taxis ($22 \approx 53 \cdot 0.5 \cdot \frac{7.5}{7} \cdot 0.8$).

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.[10]

Demographic information presented in Table 1 shows that the study reached a broad sample, suggesting 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 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.[11] However, the sample is not perfectly representative. More males (981) than females (444) were surveyed, likely because men spend more time away from home in urban Kenya. 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 variation in beliefs to estimate VSL. There are four arms: a pure control, control, and two treatments. The treatments were presented with information about the mortality risk of motorcycles and the life-saving effectiveness of helmets, while the control arms received no such 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 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. These observations are excluded from VSL estimates since beliefs were not elicited. In practice, and reassuringly, there are no differences in demand between the pure control and control.

The treatments presented different studies of helmet effectiveness. Respondents first re-

[10]Surveyors reported recruitment rates and the share of those screened out due to helmet access to the field manager daily to verify that helmet access was rare. The field manager reported that under 10% of those screened out reported helmet access, less than 1% of approached passengers.

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

ceived an estimate of their unhelmeted mortality risk over the 5 year lifespan of a helmet.[12] They were then presented with one of two studies. Those in a "low treatment" group were presented the results of Liu et al. (2008) which estimates, via a meta-analysis, that helmets reduce one's likelihood of dying by 42%. Respondents assigned to a "high treatment" arm were presented with an estimate from Ouellet and Kasantikul (2006) that helmets reduce fatality risk by about 70% in Thailand. Surveyors followed standard scripts to present the information, so only the study and results vary across arms. The sources of information and the fact that the studies of helmet efficacy were conducted outside of Kenya were disclosed.

Both studies of helmet efficacy are from peer reviewed publications. There is a strong consensus that motorcycle helmets are effective, but there is uncertainty about exactly how well they work, particularly across contexts and helmet types. The 42% figure presented in Liu et al. (2008) is often presented by the UN and NGOs operating in Kenya, motivating its use in this study. But their meta analysis primarily considers studies of high-income settings where drivings speeds and helmet quality differ. Ouellet and Kasantikul (2006) present estimates from Thailand, which may be more similar in driving and helmet environment. It is unclear which estimate is more representative of the efficacy of helmets in Kenya, meaning neither arm was given any information that I believed was inaccurate.[13]

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

[12] I calculated per trip risk for the average Kenyan from NTSA data, then estimated 5 year risks based on the respondent's ridership. Mortality rates were obtained by dividing deaths the year before the study across Kenya by 22 million (trips per day in Kenya) times 365. I combined the per trip estimate with the respondent's 5 year ridership, obtained by scaling trips/week.

[13] An ethics appendix details steps taken by the study to ensure that no respondents were harmed.

ensure that cash on hand did not determine demand.

The maximum draw was based on assessments by an NGO and the helmet manufacturer that most valuations would fall below KsH 600. Ethical concerns were also raised that if draws significantly exceeded market prices, consumers may infer that helmets were expensive which could harm their diffusion. The manufacturer sold the helmets used in this study at a price of Ksh 580, 15% of weekly median wages. A limitation of the willingness to accept mechanism is that one's true valuation is not the unique weakly dominant bid if an agent knows their valuation exceeds the maximum draw. To hedge against the risk of setting the maximum too low, and to avoid anchoring, the BDM script did not mention the upper bound or helmet value. In principle, one could infer the maximum draw from the consent form, which disclosed possible gifts but did not tie them to the game. The consent was presented well before the BDM mechanism, making it unlikely that information would be recalled. In piloting, surveyors verified that respondents could not state the value. The maximum was revealed when requested, which happened once.

In practice, helmet valuations often exceed the wholesale price.[14] Appendix Figure A2 demonstrates that there is no unusual behavior in bids near Ksh 600. In addition, agents in the low treatment arm are less likely to receive a helmet ($p < 0.1$ and $p < .05$ if baseline risk is above 1 in 100,000), validating that valuations affect real outcomes. I also show in a robustness check that results are robust to using an IV probit estimator that is not influenced by variation in valuations above the maximum draw. Thus the mechanism appears to accurately capture willingness to pay.

2.5 Randomization

Respondents were assigned to the information arms using a random number drawn in SurveyCTO. In wave one, respondents were assigned to the pure control with probability 0.1 and each of the other groups with probability 0.3. In the second wave, the pure control was

[14]High demand is reconciled with low VSL because consumers helmets for reasons other than their life saving potential. Respondents frequently cited protection from non-fatal injuries, resulting in high hospital bills and missed wages, as a leading attribute during piloting.

eliminated (since demand was not significantly different across the control and pure control in wave 1) and respondents were assigned to the remaining arms with equal likelihood.

3 Model and identification

This section presents a simple model of demand for helmets and illustrates the method used for estimating demand for non-market goods, showing how it identifies 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 $i \in \{1, \ldots, \infty\}$. 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 when beliefs are proxied by empirical risk. 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 while wearing a helmet in a motorcycle accident that would be fatal without a helmet given by

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

where $\mathcal{I}_i$ denotes the individual's information set and D is a Bernoulli random variable equal to 1 if the agent dies. Beliefs are over the unknown parameter of this 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

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

Suppose that the consumer receives a signal that the estimated efficacy of helmets is $\theta_H \sim Binomial\left(a_{EH} + b_{EH}, a_{EH}/(a_{EH} + b_{EH})\right)$. The Binomial likelihood captures the fact that studies of helmet efficacy report the number of deaths given a 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 posteriors about the efficacy of helmets are

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

with expected value

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

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, so 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 beliefs. If agents learn through their experiences or their social network then beliefs will likely vary from empirical estimates, violating rational expectations. Learning from one's experiences is subject to survivor bias. Social networks information is prone to small sample biases. In addition, a higher likelihood of learning about fatal trips would cause agents to overestimate risk.

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

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

where A is a Bernoulli random variable equal to 1 if a trip ends in an accident. Beliefs are again over the unknown parameter of this Bernoulli distribution.

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

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

$[1 - Pr(A = 1|\mathcal{I}_i)]^{n_i}$ 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 priors about $Pr(A = 1|\mathcal{I}_i)$ are non-degenerate, one must integrate over beliefs to obtain equation 1.

I assume that helmeted individuals involved in an accident that would otherwise be fatal are deterred from continuing to use motorcycles.[15] 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

$$r_{ih}(z_i) = \begin{cases} H_{i0} \cdot r_{iu}, & z_i = 0 \\ H_{i1} \cdot r_{iu}, & z_i = 1 \end{cases} \tag{2}$$

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, health, and demographics. Denote their flow utility of consumption by $u(c_i; x_i)$ and denote the expected utility from not being alive (which may be positive, e.g. if agents believe in an afterlife) by $u_d(x_i)$. The agent's expected utility from purchasing a helmet is

$$U_{ih}(z_i) = \zeta_h + [1 - r_{ih}(z_i)] \cdot u_a(x_i) - p_i \cdot u'(c_i; x_i) + r_{ih}(z_i) \cdot u_d(x_i) + \epsilon_{ih} \tag{3}$$

ζ_h captures average utility from characteristics of helmets other than mortality risk reduction (e.g. injury protection, comfort), and ϵ_{ih} denotes idiosyncratic variation in utility.

[15] Absent this assumption, $r_{ih0} = \frac{H_{i0} \cdot r_{in}}{1 - H_{i0} \cdot r_{in}}$. Results are similar since suffering two accidents is unlikely.

Without purchasing a helmet, expected utility is

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

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

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

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,

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

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

$$VSL_i \equiv \frac{dp_i}{d\Delta r_i} = \left(\frac{\partial U_{i,h-u}}{\partial \Delta r_i}\right) \Big/ \left(\frac{\partial U_{i,h-u}}{\partial p_i}\right) = \frac{u_a(x_i) - u_d(x_i)}{u'(c_i; x_i)} \tag{4}$$

This predicts that VSL will increase with income. Theory suggests that $u_a(x_i)$ will increase and $u'(c_i; x_i)$ will fall as agents become richer. $u'(c_i; x_i)$ is likely to be particularly important as estimates often indicate it varies non-linearly with income (Havránek, 2015). This reflects a lower opportunity cost of funds, so VSLs will be higher even if agents obtain no more utility from averting a death. Expected utility from not being alive, $u_d(x_i)$, may also vary substantially across contexts, especially with beliefs about an afterlife.

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 across agents. Let Δr_i^* be a statistical estimate of the likelihood a helmet will save one's life. A common set of identifying assumptions in the VSL literature (but not this paper) is

$$\text{Full Information Rational Expectations:} \quad \Delta r_i = \Delta r_i^* \quad \forall i \tag{5a}$$

$$\text{Exogeneity and Logit Errors:} \quad \epsilon_{ih}, \epsilon_{iu} \sim_{iid} EV1 \tag{5b}$$

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

$$Pr(y_i = 1) = \Lambda\left(\zeta_h + \beta \cdot \Delta r_i^* - \alpha \cdot p_i\right) \tag{6}$$

Λ is the Logistic CDF and so α, β and thus VSL, which is $\frac{\beta}{\alpha}$, are identified.[16]

Assumption 5a is violated if $Pr(\Delta r_i \neq \Delta r_i^*) > 0$, meaning agents hold beliefs about risk that diverge from statistical estimates, even if those estimates match beliefs in expectation. This assumption is plausible when consumers have the same information as the researcher, but fails if they possess private information (e.g. personal experiences) or lack access to the data used by the researcher. 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 noisily measured.[17]

Assumption 5b is violated if mortality risk is endogenously related to unobserved determinants of utility. For example, frequent motorcycle riders face higher risk but may also experience greater disutility from helmet discomfort. Some studies use instruments to relax this assumption but still rely on rational expectations (e.g. Ito and Zhang, 2020).

3.2 Identification using subjective belief data

This study estimates VSL without assuming rational expectations or exogenous mortality risk by using an instrument that shifts agents' subjective beliefs. The model shows that this is an appropriate way to identify VSL since presenting agents with information about

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

[17] 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."

helmet efficacy will change the expected utility they obtain from the life saving potential of a helmet (unless their priors exactly 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

$$\text{Exclusion Restriction:} \quad Cov\left(z_i, \epsilon_i\right) = 0 \tag{7a}$$

$$\text{Instrument Relevance:} \quad Cov(\Delta r_i, z_i) \neq 0 \tag{7b}$$

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

$$\begin{aligned}
\epsilon_{iu} &= \zeta_h + \beta \cdot \Delta r_i(z_i) - \alpha \cdot v_i + \epsilon_{ih} \\
\alpha v_i &= \zeta_h + \beta \cdot \Delta r_i(z_i) + \epsilon_i \\
v_i &= \frac{1}{\alpha}\zeta_h + \frac{\beta}{\alpha} \cdot \Delta r_i(z_i) + \frac{1}{\alpha}\epsilon_i = \zeta_h' + VSL \cdot \Delta r_i(z_i) + \epsilon_i'
\end{aligned} \tag{8}$$

VSL is identified from data on v_i and (potentially misreported) beliefs $\Delta r_i(z_i)$, using information exposure z_i to instrument for beliefs (ζ_h' is an intercept). When z_i is randomized, then denoting measurement error free beliefs by Δr_i^t, assumption 7a is equivalent to

$$\text{Classical Measurement Error:} \quad \mathbb{E}[\Delta r_i - \Delta r_i^t] = 0 \tag{7a$'$}$$

This holds since z_i only contains information about Δr_i and (by randomization) is independent of other drivers of v_i. So exclusion holds if misreporting is uncorrelated with z_i.

The model shows that relevance typically holds under Bayesian learning, but estimates are consistent as long as information changes beliefs (even if agents learn in non-standard ways). Relevance may be verified empirically by testing if $Cov(\Delta r_i, z_i) \neq 0$.

The linearity of VSL in mortality risk follows from considering marginal reductions in risk. For large changes, the relationship would likely be convex since the marginal utility of consumption would fall with transfers, so applying VSL may underestimate welfare gains.

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. Or to study preferences for privacy, r_i may be the risk of a privacy breach reduced by security software.

4 Data and empirical specification

4.1 Data

I use data from 1,571 surveys collected 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. The second included 650.

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 in piloting. Details of the final approach are presented in Appendix A and the full instrument is available on the AEA RCT Registry.[18] The survey elicits priors about per trip mortality risk, 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 unit interval (such as less than 1 in 10 million, 1 in 10 million to 1 in 1 million, up to greater than 1 in 10). Then respondents were then asked to list a more precise belief within the interval. This two-step approach reduced rounding in piloting.[19]

[18] The instrument is appended to the end of the PAP amendment.

[19] There is likely still rounding, especially of high beliefs. Such rounding likely results in classical error, but I test for bias from non-classical error. Results are robust to including fixed effects for the interval selected.

Table 1 summarizes demographics of the sample and demonstrates balance across treatments. Appendix Table A1 is similar, but it examines motorcycle taxi use and beliefs. These outcomes are generally balanced across arms, although there is some imbalance between the pure control (which had a smaller number of respondents) and other arms.

4.2 Experimental VSL estimation

The primary estimate of VSL is obtained via the two-stage least squares model

$$\begin{aligned} v_i &= \zeta_h' + VSL\Delta r_i + X_i'\gamma_0 + \gamma_1 r_{0,i} + \epsilon_i' \\ \Delta r_i &= Z_i'\pi + X_i'\pi_c + \pi_r r_{0,i} + \nu_i \end{aligned} \tag{10}$$

where v_i is the respondent's willingness to pay for a helmet, Δr_i is their perceived reduction in mortality risk from a helmet, X_i denotes controls, and $r_{0,i}$ is the respondent's baseline belief about their unhelmeted mortality risk. I report results with two sets of 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 absorb heterogeneity in priors, greatly improving power. Intuitively, if there is variation in the perceived riskiness of motorcycles, then Δr_i will vary both due to beliefs about risk and the efficacy of helmets. The interacted instruments capture this, so the first stage more accurately predicts beliefs. This set of instruments is similar to that studied in Abadie et al. (2023), which shows that similar instruments improve asymptotic mean squared error.[20]

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.[21]

I follow the pre-analysis plan and report results over two samples. First, I use data from

[20] I pre-specified two sets of IVs, but the original pre-analysis plan (PAP) registered a slightly different version of the interacted IV using n_i, motorcycle trips/week, instead of $r_{0,i}$. This inaccurately assumed n_i predicts $r_{0,i}$. An amendment before wave 2 specified $r_{0,i}$.

[21] Due to an error, the initial PAP only listed the demographic variables. The PAP amendment specified the full set of potential controls and enumerator FEs. Results are similar if they are excluded.

the control and both treatments. Second, I restrict the sample to treated respondents to help rule out confounds. The primary tables report homoskedastic standard errors 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 compare the experimental VSL estimates to two traditional approaches used to estimate the parameter. First, I estimate

$$v_i = \zeta_h + VSL\Delta r_i^* + X_i'\gamma_0 + \epsilon_i \tag{11}$$

where Δr_i^* is the empirical likelihood that a helmet will save the respondent's life, estimated from ridership and an estimate of helmet efficacy. This approach is similar to León and Miguel (2017) and Ito and Zhang (2020) which estimate random coefficient logit models under rational expectations. 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

$$VSL_i = \frac{v_i}{\Delta r_i^*} \tag{12}$$

This assumes rational expectations and that agents do not receive utility from any characteristics of helmets other than safety. This specification follows Berry et al. (2020) and is similar to Kremer et al. (2011). The method is typically used in the context of health or safety products, where assumptions are more plausible.

These approaches typically do not separate willingness to pay for non-fatal illness/injury prevention from mortality risk reduction. Therefore authors often interpret estimates as upper bounds on VSL (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 a fatal motorcycle accident fall within plausible ranges and are strongly correlated across measures, supporting their use in estimating VSL. The mean reported 5-year mortality risk is 0.034, with a median of 0.001.[22] For comparison, the estimated median empirical risk is about 1 in 4,000, though deaths may be under-reported. Beliefs about deaths per 10,000 passengers and one's own risk are highly correlated ($R^2 > 0.1$), though respondents view themselves as less at risk than average.

Beliefs about 1-year and 5-year mortality risk are also consistent ($R^2 = 0.57$). over 60% of respondents report 5-year risk estimates between 4 and 6 times their 1-year estimate, and less than 1% gave logically inconsistent answers. Respondents' 5-year risk beliefs are also aligned with per-trip fatality risk assessments ($R^2 = 0.2$).

First stage: Effect of information on beliefs

Table 2 demonstrates that randomized information exposure had a statistically significant effect on the agents' beliefs that a motorcycle helmet will save their life.

Respondents were first presented with estimates of their unhelmeted mortality risk constructed using data from the NTSA. Anecdotal reports indicate mistrust in the NTSA and, consistent with mistrust, respondents reported no change in beliefs in response to this information (columns 1 and 2). There is not a significant difference between either treatment arm, which received identical information, and the control group. This suggests agents did not feel compelled by experimenter demand effects to misreport changes in beliefs.

Respondents did update beliefs about the effectiveness of helmets based on the studies presented in the treatments, the primary variation used by the interacted instrument. 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 effective-

[22] Results are robust to excluding implausible responses or winsorizing beliefs (Appendix Table A2).

ness reported in the control group was 79%, and many respondents stated beliefs over 90%. Exposure to the 70% (high) and 42% (low) effectiveness estimates reduced beliefs to 75% and 65%, respectively (columns 3–4), with the treatment gap significant at the 1% level.[23]

The change in beliefs about helmet effectiveness was sufficient to shift the perceived probability that a helmet will save a respondent's life (the product of the prior two variables). The dependent variable in Columns 5 - 6 of Table 2 equals 10,000 times beliefs. This is 80 units (36%) lower on average in the low treatment versus control group, significant at the 1% level. The high treatment group does not differ significantly from control (because the signal is close to priors), but point estimates move in the expected direction.

These results establish that information affected beliefs, confirming that treatment assignment is a relevant instrument for perceived reductions in mortality risk.

The value of a statistical life

Primary estimates of the value of a statistical life are reported in Table 3. The preferred estimate in column 1 is \$224, indicating low demand for safety. A weak-instrument robust 95% confidence set excludes values below \$34 and above \$429. Estimates are consistent across specifications, both with and without controls, and across instrument sets.

The estimates pass a battery of robustness checks. Panel b shows that excluding control observations does not materially affect results, though estimates are somewhat larger. Despite reduced power, values above \$1,038 are still rejected at the 95% level. Panel c similarly shows that leaving any experimental arm out does not meaningfully alter estimates.

This method for estimating VSL does not require experimental variation if a valid observational instrument for subjective beliefs exists. Leveraging this, I instrument for risk using a plausibly exogenous shifter of beliefs in column 4: an indicator for knowing a motorcycle accident victim (conditioning on taxi terminal fixed effects). Estimates are remarkably similar to those obtained from random variation, with the mean VSL estimate in column 4 at \$560. In fact, a J-test fails to reject the equality of estimates when both this variable and treatment assignment are used as instruments.

[23]Under Bayesian updating one would not expect a 30 p.p. gap in beliefs between the treatment arms since beliefs are a weighted average of priors and the signal. Estimates both imply a weight of about 60% on priors.

Appendix Table A2 demonstrates that results hold if beliefs about the probability that a helmet will save one's life are winsorized. Appendix Table A3 shows that the results are robust to changes in planned future ridership (panel a) or subjective beliefs about helmet lifespan (panel b). Panel c in the table suggests that results are not driven by selection into high ridership because of low demand for safety.[24] Appendix Table A4 shows that IV probit results regressing receipt of a helmet on price (the BDM draw) and perceived risk reduction (instrumented with treatment status) also yields similar results, validating that results are not influenced by inaccurate valuations above the maximum BDM draw.

Although demand for safety in this setting appears low, economic theory suggests these estimates are consistent with much higher VSLs in richer settings, where the opportunity cost of funds is far lower. Comparing results from this paper to estimates across contexts reveals that this study's estimates are of a similar order of magnitude to those from similar income levels, but much smaller than those from wealthier samples. Some differences may be due to methodology, but this helps to confirm the importance of income.

Figure 1 plots point estimates and standard errors from all LMIC revealed preference estimates of VSL that I am aware of. An estimate from Greenberg et al. (2021) is also presented, which examines US soldiers using related methods, for comparison. Estimates from this paper are lower than most existing values, but more 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 near $0 and just over $4,000 in Ghana. Ito and Zhang (2020) and León and Miguel (2017) find much higher VSLs among wealthier populations within LMICs.

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 differences in VSL. Under a CRRA utility function, the gap between this study's estimate and the VSL of $700,000 in León and Miguel (2017) implies

[24]Estimates exclude observations from the second day of data collection, as specified in the PAP amendment, because in some cases motorcycle drivers were pretending to be passengers to obtain a free helmet and submitting false surveys. Estimates are similar but less precise if they are included (95% CI -$159 to 241).

a coefficient of relative risk aversion of $\theta < 3$.[25] This falls solidly in the range of empirical estimates: Havránek (2015) reports a mean elasticity of inter-temporal substitution of $1/3$ in a meta-analysis, equivalent to $\theta = 3$. While CRRA utility may not hold exactly, and methodological differences may drive gaps in results, this exercise demonstrates that large cross-country variation in VSL can readily be rationalized by differences in the marginal utility of consumption. This is consistent with Hall and Jones (2007), which finds that health investment as a share of consumption rises sharply with income.

Is an average VSL below annual income plausible?

The average VSL estimated in this study is only a fraction of annual income. Is such low demand for safety plausible?

An important insight for interpreting this result is that VSL is not equivalent to the present value of lifetime consumption, especially if individuals believe in an afterlife (like 87% of Kenyans surveyed)[26] or have fatalistic views (León and Miguel, 2017). Estimates also align with a consistent finding that demand for effective preventive health products is often low in LMICs (Dupas and Miguel, 2017). These investments offer a useful point of comparison since they involve similar tradeoffs between modest mortality risks and cost.

A number of studies illustrate low demand. Cohen and Dupas (2010) show that only 40% of pregnant women in Kenya pay \$0.60 for an insecticide-treated bed net that all adopt for free. Malaria nets save an estimated 5.6 lives per 1,000 users annually, implying a cost per life saved below \$110 at \$0.60 (Pryce et al., 2018). Ashraf et al. (2010) find that demand for water treatment, which prevents a leading form of child mortality in diarrhea, drops 30 percentage points when price rise from \$0.10 to \$0.25. Uptake of deworming pills falls by 60 percentage points when costs rise by \$0.30 (Kremer and Miguel, 2007). And Banerjee et al. (2010) show that offering in-kind incentives worth \$4 to vaccinate children (at no cost) increases vaccination rates from 18% to 39%. Estimates suggest vaccines reduce

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

[26] Gallup International Association. (2022). *More Prone to Believe in God than Identify as Religious. More Likely to Believe in Heaven than in Hell.*

child mortality by at least 0.5 percentage points in India (Kumar, 2024). Moreover, Dupas and Miguel (2017) document that demand for such products typically remains well below 100% when capital or psychological constraints are alleviated, consistent with a low VSL.

When does economic theory support the use of VSL estimates in poor countries?

These results should not be interpreted as a call to eliminate health funding to LMICs because the use of the estimates is not theoretically supported to value all health investments. The estimates are only valid for decisions where a population faces a mortality risk-consumption trade-off. This informs many important decisions such as those discussed in the next section. But it does not apply when funds are earmarked for health.

To illustrate this, suppose that there are otherwise homogeneous high and low-income consumers indexed by h, ℓ. As in section 3, let α denote the marginal utility of consumption and β expected utility from averting death. Suppose a high-income government distributes N vaccines to low-income consumers, each of which saves a life with probability p. The cost is $c \cdot N$ paid for by N high-income consumers. The global welfare gain is

$$W = N \cdot p \cdot \beta_\ell - N \cdot c \cdot \alpha_h$$

This is positive if $p \cdot VSL_\ell \cdot \frac{\alpha_\ell}{\alpha_h} > c$. Thus, evaluating policies by estimating lives saved, valuing them at the low-income country's VSL, and comparing benefits to costs will understate welfare gains by $\frac{\alpha_\ell}{\alpha_h}$. This may be orders of magnitude. For instance, with the CRRA value of 2.91 that fits VSL differences described above, the marginal utility of a dollar would be about 4,000 times less valuable in the United States (PPP GDP/capita $83,000) versus this sample (median income $4,750). Intuitively, low-income consumers have low VSLs due to a high opportunity cost of funds. If high-income consumers pay program costs, low-income consumers face no consumption tradeoff.

One example where this distinction matters 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 income levels. Agrawal et al. (2023) argue that more equitable

COVID-19 vaccine allocation may have saved 670,000 lives but reduced welfare due to VSL differences. But LMICs relied on donated vaccines (Ahuja et al., 2021). Hence, these studies likely underestimate the welfare gains of lives saved in poor countries.

In addition, applying VSL to value large changes in risk – such as HIV treatment to prevent near certain death – may be inaccurate. Preferences for mortality risk may be convex if transfers large enough to offset major risks affect the marginal value of consumption. Thus, the VSL estimates in this paper should be applied only to policies involving modest mortality risks where consumers face a consumption-risk tradeoff.

Welfare implications of the estimated VSL

The welfare implications of the VSL estimates presented depend on the values that are currently used in benefit-cost analysis. If different VSL values are considered, then using these estimates could yield welfare gains by aligning policy with preferences.

I selected the five most recent benefit-cost analyses in Kenya from Google Scholar that use VSL to assess how these results affect policy conclusions.[27] I re-calculate benefit-cost ratios (BCRs) using the preferred VSL estimate and compare them to the original findings. Appendix Figure A3 shows that BCRs fall sharply: on average by over 99%, and in four cases the BCR falls from above to below 1, reversing the original policy conclusion.

The results also suggest that some international development assistance may be misallocated. GiveWell, an NGO that has matched over $1 billion to charities, weights averting the death of someone aged 15-49 at 104 times doubling their consumption for a year,[28] directly affecting which charities are recommended. This paper suggests that these weights may under-value the benefits of consumption gains by orders of magnitude, suggesting that allocating more funding towards programs like cash transfers could enhance welfare.

Heterogeneity in the value of a statistical life

If VSL is heterogeneous, estimates based on a selected population may differ from the average VSL relevant for targeted policies. This paper focuses on urban commuters, a

[27] Accessed June 2024, including Babagoli et al. (2022) (menstrual cups/sanitary pads), Mwai et al. (2023) (primary health), Hamze et al. (2017) (cleft lip/palate repair), and Oyugi et al. (2023) (maternity care).

[28] Based on 2023 update. Weights are influenced by a stated preference VSL (Redfern et al., 2019).

group of interest for transportation policy. To assess broader applicability, I examine VSL heterogeneity on observables in Table 4. Results align closely with theoretical predictions, but the degree of heterogeneity across individuals in the sample is modest, suggesting the conclusion (demand for safety is low) is unlikely to change across similar populations.

Column 1 shows that VSL is similar below and above the median age. This is consistent with Aldy and Viscusi (2007), which find that VSL peaks near age 40.

Columns 2 and 3 show a highly statistically significant positive relationship between VSL and income, matching theory. Column 2 reports heterogeneity with respect to asinh(wages), demeaned so the first coefficient captures average VSL.[29] Estimates indicate that a 1% increase in wages is associated with an increase in VSL near $4. This implies a back of the envelope income-elasticity of 1.14. Column 3 shows agents with above median wages have a VSL that is higher by $500.

Column 4 provides evidence that VSL is also increasing with health. There is not significant heterogeneity with respect to having children, digit span recall (a cognitive test of memory), years of education, or gender. This implies that the fact the study sampled more men than women is not likely to substantially influence estimates.

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

The primary identifying assumption is that the gap between agents' true subjective beliefs and their reported beliefs reflects only classical measurement error. True subjective beliefs need not match the empirical truth, and research shows that agents decision-relevant beliefs can be influenced by heuristics like rounding (Lacetera et al., 2012). Tests reported below support the view that errors are classical, suggesting that VSL estimates are consistent.

The experiment included two treatments of different intensity so that VSL may be estimated at multiple points of the belief distribution. Estimates should be invariant to the experimental arms included under classical error. However, if beliefs were misreported – unless misstatement was proportional across arms – estimates would differ. For example,

[29] I set the wages of the unemployed to 0 and use asinh since unemployment is often involuntary. Results are similar, but less precise, dropping the unemployed and using log(wage).

if lower reported helmet efficacy in the low treatment arm reflected experimenter demand effects, estimates using the low treatment and control groups would be smaller (since the denominator would be over-estimated) than those using the high treatment and control. Estimates are similar across all combinations of arms (Table 3), helping rule out this concern.

The treatment only and interacted instruments facilitate a second test. In short, the interacted IV places greater weight on individuals with higher priors about motorcycle risk. If reported beliefs were subject to non-classical measurement error correlated with true beliefs, the IVs would yield different estimates. Appendix C presents a formal argument, based on the literature on local average treatment effects (e.g. Abadie et al., 2023; Mogstad et al., 2018), and shows that this analysis can test for common patterns of misreporting.

Intuitively, willingness to pay for a helmet is a function of an agent's true belief about motorcycle risks, $r_{0,i}$, helmet effectiveness, and their VSL. The treatment only estimates are proportional to the covariance between willingness to pay and treatment assignment, so a weighted average of VSLs proportional to $r_{0,i}$ is identified. With the interacted instrument, first stage predictions also depend on elicited $r_{0,i}$. This leads to a weighted average of VSLs proportional to $r_{0,i}^2$ if errors are classical. If beliefs are systematically mismeasured, weights will be a function of the correlation between $r_{0,i}$ and error with this instrument, whereas treatment only weights will remain proportional to $r_{0,i}$. Thus, the IVs will generally produce different estimates, unless errors are orthogonal to true beliefs.

For instance, some agents report very high risks of riding motorcycles, resulting in average reports much larger than empirical estimates. Are they overstating their belief? This would result in the interacted IV producing a lower estimate, as long as those reporting very high beliefs have above average true expectations. The appendix argues that typical forms of rounding and demand effects would also cause the IVs to diverge.[30] Yet empirically, VSL estimates are similar across instrument sets, and a Hansen J-test fails to reject their equality ($p > 0.5$), suggesting that significant bias from misreported beliefs is unlikely.

Experimenter demand effects are further ruled out by the reduction in willingness to

[30]Appendix Figure A4 illustrates this in the case of rounding.

pay in the low treatment arm, providing revealed preference evidence of belief updating. And helmet demand is similar in the pure control and control arms, despite the fact that the control was asked about the risks of motorcycles, which could plausibly generate demand effects. Lastly, as noted above, VSL estimates based on quasi-exogenous variation from accident victim exposure are statistically indistinguishable – further reinforcing this point.

To further address concerns about bias from difficulty articulating small probabilities, I proxy for perceived helmet risk reduction by combining beliefs about the number of motorcycle deaths per 10,000 users over 5 years with reported helmet efficacy. This avoids relying on small probability elicitations. I estimate a VSL of \$151, which is close to (and statistically indistinguishable from) the preferred estimate. The lack of VSL heterogeneity with respect to education or digit span recall is also inconsistent with this concern.

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 likely biased.

This study captures both agents' subjective beliefs and their empirical mortality risk, allowing for tests of rational expectations. The two variables are uncorrelated ($\rho = -0.0008$) and respondents report average beliefs above empirical estimates, suggesting a violation of rational expectations. Measurement error could obscure a relationship, so I instrument for final beliefs using another measure (deaths/10,000 users), which addresses bias from independent classical errors (Gillen et al., 2019). This yields a statistically significant positive relationship, but the coefficient is below 0.001. I reject equality to 1 (rational expectations) with 99.9% confidence. However, as 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 asked respondents to list the information sources they used to form their beliefs. Figure 2 plots the responses. The most common source is own experience (79%), followed by family members (48%) and social media (38%). Under a third of respondents reported consulting more objective media or government sources. These information sources are consistent with deviations from rational expectations since they depend on private information and are prone to bias.

Second, I regress beliefs on (1) variables that shift empirical risk and (2) variables that should influence beliefs (but not necessarily true risk) if reported information sources are accurate. Specifically, I use trip volume and duration as empirical shifters, since more frequent motorcycle use mechanically increases risk, particularly conditional on taxi terminal fixed effects. Consistent with this, a 1% increase in ridership is associated with a 0.13 percentage point increase in the likelihood of reporting a prior accident in the data ($p < .01$). To test if agents form beliefs based on personal or social network experiences, I consider indicators for having suffered a motorcycle accident or knowing someone who has.

If rational expectations held, 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. In contrast, accident history – personal or in one's social network – is strongly predictive of beliefs. These patterns suggest a violation of rational expectations.

One concern is that accident history may be driven by risk heterogeneity observed to the respondent but not the econometrician, in which case rational expectations could hold under a more complete model of risk. At a minimum, this demonstrates the sensitivity of results to the researcher's data. Moreover, I examine if respondents reported higher 5-year risks if it rained the day of the survey. The weather today should not affect long-run risk, but rain increases the salience of conditions when motorcycles are dangerous. Respondents reported about twice the risk if it rained, suggesting a departure from rational expectations.

I next examine if the approaches typically used to estimate VSL, which assume rational expectations, consistently estimate VSL in Table 6.[31] Estimates from traditional methods generally fall outside the experimental confidence sets, and different observational approaches yield starkly different results.

[31] One must take a stance of the efficacy of helmets. Panel a reports results using Liu et al. (2008). Panel b reports estimates using Ouellet and Kasantikul (2006). Discussion focuses on panel a.

Columns 1 - 4 report estimates of VSL obtained from Equation 11. Column 1 omits covariates; columns 2–4 use double-post LASSO for covariate selection, with column 3 adding enumerator and column 4 adding taxi terminal fixed effects. Estimates are highly sensitive to covariates, and confidence intervals are wide. The point estimate exceeds \$15,000 without controls and is negative in column 3. Columns 2 and 4 yield point estimates closer to the experimental benchmark, around \$2,700 and \$1,500, respectively. While this offers some evidence that rational expectations-based methods can recover VSL, standard errors are large and these results may be driven by a weak correlation between willingness to pay and ridership. Estimates using Equation 12 are reported in column 5. The mean estimate exceeds \$380,000.

The experimental approach relaxes rational expectations and addresses endogeneity, raising the question of which force drives differences between methods. Column 6 estimates VSL using subjective belief data, instrumenting for final beliefs with priors about the number of motorcycle deaths/10,000 passengers over 5 years. If measurement error is independent across measures, this corrects attenuation bias but not omitted variable bias. The resulting estimate, about \$110, is smaller than but not statistically different from the experimental estimate. While suggestive, this indicates that deviations from rational expectations may be more important than endogeneity in this setting, consistent with subjective beliefs being shaped by idiosyncratic experiences, such as accident realizations.

5.4 Other threats to identification

There are several threats to the experimental VSL estimates aside from misreported beliefs.

First, the treatments may have influenced beliefs about the efficacy of helmets at injury prevention, not just mortality. To limit this, the information excluded injury content and no injury-related questions were asked. If injury beliefs did shift, the VSL estimates would represent upper bounds, not affecting the conclusion that safety demand is low.

Second, VSL may be underestimated if agents did not plan to regularly use helmets. Piloting indicated that respondents believed that they would consistently use a helmet if they received one, so bias on this dimension seems unlikely (although possible). Resale is also improbable because respondents all accepted a helmet with quality tags and the box removed, which undermines secondary market value but avoids the hassle cost of disposing of waste if the recipient intends to wear it. And even if respondents privately anticipated using a helmet only 10% of the time, VSL estimates would remain well below those used in policy. However, some respondents did report a risk they would lose the helmet. Appendix Table A3 shows that estimates are similar when adjusting for anticipated duration of use.

Third, monotonicity violations could affect the interpretation of estimates as a weighted average VSL. However, in practice, only 3% of control respondents had priors below the 42% figure presented to the low treatment group, and estimates are similar when only control and low treatment observations are considered.

Could mismeasurement of helmet valuations bias results? Prior research shows that the BDM effectively recovers valuations in similar high-stakes environments (Berry et al., 2020). Moreover, a level shift in valuations would not bias VSL estimates, which are identified from differences in willingness to pay across information treatments. The similarity of IV probit estimates also provides evidence against this confound.

Finally, agents could face capital constraints. Demand was elicited via a helmet-versus-cash choice to ensure valuations are not constrained by cash on hand. While temporary liquidity needs may exist, the estimated mean VSL is only \$500 among respondents with above median income and median willingness to pay is $1/4$ weekly wages, suggesting substantial bias is unlikely. And estimates are about \$290 if the unemployed are excluded. Moreover, since recipients of aid programs likely face similar constraints, the VSL estimates should still accurately reflect the value of consumption gains, even if imperfect capital markets contribute to high consumption value.

6 Conclusion

This study introduces a novel framework for experimentally estimating demand for non-market amenities using subjective belief data and uses the approach to estimate the value of a statistical life. I leverage the fact that products often exist which affect an agent's exposure to non-market goods. This paper demonstrates that researchers may cause agents to update their beliefs about how such a product affects the amenity of interest, then examine how product demand changes to uncover willingness to pay for the amenity. This method is tractable and low-cost in many settings where exogenous variation in attribute exposure does not exist, and it avoids the need to assume that agents' beliefs satisfy rational expectations, which appears to be violated in this setting.

I estimate low demand for mortality risk reduction in urban Kenya. This is consistent with a common finding that uptake of preventative health products is low in LMICs. Development economists have long sought to explain this finding, with no clear consensus explanation. It is certainly possible that a feature of decision making may deflate demand for preventative health and bias estimates of VSL in this paper. However, these results suggest another explanation: perhaps low demand for safety simply reflects a high marginal utility of consumption among these low income households.

These results suggest scope to improve welfare by better aligning policy decisions with recipients' preferences. Highlighting potential misallocation, a major aid organization underweights consumption gains by orders of magnitude relative to the level implied by these estimates. This has direct welfare consequences, as funds are allocated to preventative health programs at the expense of cash transfers. This does not mean that health programs are not valuable. Rather, it suggests that recipients may value consumption so strongly that the benefits of higher incomes can exceed those of some health investments. However, theory supports applying VSL only when a direct health-consumption tradeoff is faced, so these findings do not support reallocating health funds towards richer populations.

Finally, this paper suggests that VSL is likely to vary substantially across populations due to factors like wealth and religious beliefs. However, methodological differences between estimates make it challenging to study the sources of VSL heterogeneity. It may therefore be valuable to apply the method in this paper in additional contexts to provide insight into the determinants of demand for mortality risk reduction.

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). 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. *Annals of Mathematical Statistics 20*(1).

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). Estimating the Value of a Statistical Life: The Importance of Omitted Variables and Publication Bias. *American Economic Review 94*(2).

Ashraf, N., J. Berry, and J. M. Shapiro (2010). Can Higher Prices Stimulate Product Use? Evidence from a Field Experiment in Zambia. *American Economic Review 100*(5).

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.

Banerjee, A. V., E. Duflo, R. Glennerster, and D. Kothari (2010, 6). Improving immunisation coverage in rural India: clustered randomised controlled evaluation of immunisation campaigns with and without incentives. *BMJ (Clinical research ed.) 340*(7759), 1291.

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). Inference on Treatment Effects after Selection among High-Dimensional Controls. *The Review of Economic Studies 81*(2).

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

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

Cohen, J. and P. Dupas (2010). Free distribution or cost-sharing? Evidence from a randomized malaria prevention experiment. *Quarterly Journal of Economics 125*(1), 1–45.

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

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

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. *NBER Working Paper 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.

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

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

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

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

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

Kremer, M. and E. Miguel (2007, 8). The Illusion of Sustainability. *The Quarterly Journal of Economics 122*(3), 1007–1065.

Kumar, S. (2024). The Effects of Childhood Immunization Program on Health and Education: Micro-Evidence from India.

Lacetera, N., D. G. Pope, and J. R. Sydnor (2012, 8). Heuristic Thinking and Limited Attention in the Car Market. *American Economic Review 102*(5), 2206–36.

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.

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). Using Instrumental Variables for Inference About Policy Relevant Treatment Parameters. *Econometrica 86*(5).

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

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

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

Pryce, J., M. Richardson, and C. Lengeler (2018, 11). Insecticide-treated nets for preventing malaria. *The Cochrane Database of Systematic Reviews 2018*(11), CD000363.

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). 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).

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

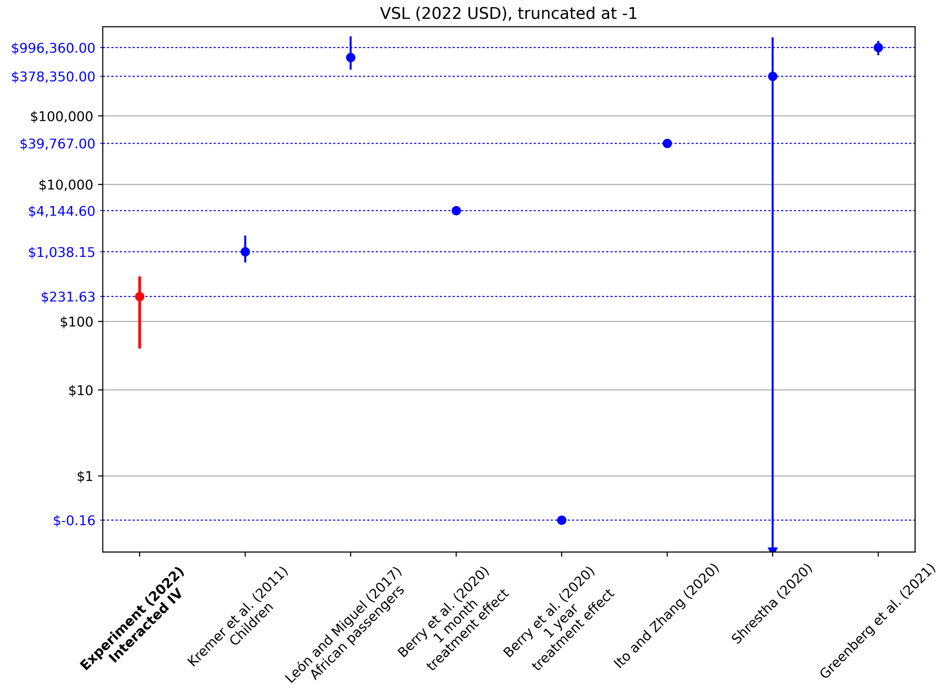


Figure 1 plots revealed-preference VSL estimates and (where available) 95% confidence intervals from this paper, Kremer et al. (2011, rural Kenya), León and Miguel (2017, wealthy international travelers in Sierra Leone), Berry et al. (2020, rural Ghana), Ito and Zhang (2020, urban China), Shrestha (2020, potential Nepalese migrants), and Greenberg et al. (2021, American soldiers). 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.

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

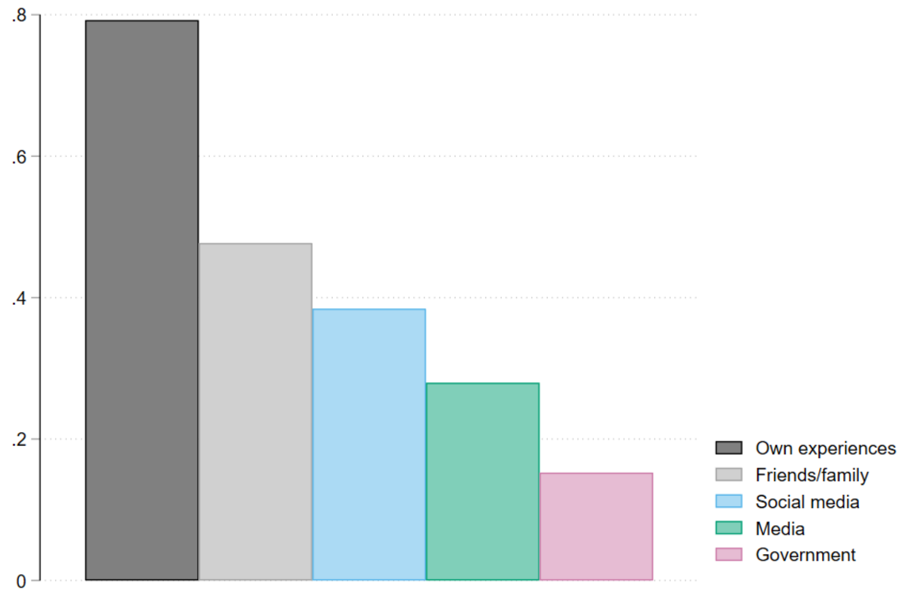


Figure 2 plots the information sources that respondents reported using to form beliefs about the mortality risk of motorcycles. Respondents were able to select multiple options, so the columns do not add to 1.

Table 1: Summary statistics and balance: Demographics

	(1) Control	(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 ($1000s)	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]	6.166	-0.636	-0.925	-1.422*	-0.491
/wage today	[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)
Primary school	0.964	-0.035	0.007	-0.022	-0.029**
complete	[0.186]	(0.026)	(0.013)	(0.013)	(0.013)
Secondary school	0.722	-0.047	0.024	-0.008	-0.031
complete	[0.449]	(0.058)	(0.029)	(0.029)	(0.028)
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	
Joint p-value		0.743	0.183	0.193	0.271

Standard deviations in brackets. Standard errors in parenthesis. * $p < .10$, ** $p < .05$, *** $p < .01$

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. The joint p-value reported at the bottom of each table reflects an F-test of joint orthogonality estimated by regressing treatment status on each of the covariates, imputing missing values with the median to avoid dropping the observations. The observation counts in columns 2-4 reflect the the size of the treatment group compared to the control. 35 observations for which surveyors reported motorcycle taxi drivers pretending to be passengers are excluded.

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 product of the prior two variables. 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) Interacted	(3) Treatment only	(4) Know victim
VSL	223.90	215.23	347.26	559.99
	(97.88)	(98.38)	(259.49)	(225.82)
Cragg-Donald F-statistic	40.75	40.12	10.99	16.24
Weak IV Robust Confidence Set	[34.22, 428.34]	[24.10, 420.73]	[-160.17, 1,066.93]	[193.40, 1,230.60]
Inversion test	CLR	CLR	CLR	AR
Observations	1,425	1,425	1,425	1,425
Controls	LASSO	BL Risk	BL Risk	None
Fixed effects	Enumerator	Enumerator	Enumerator	Terminal

Panel B: Treated respondents only

	(1) Interacted	(2) Interacted	(3) Treatment only
VSL	394.52	396.03	774.94
	(257.04)	(253.28)	(478.06)
F-statistic	16.07	16.84	11.24
Confidence Set	[-88.12, 1,038.34]	[-80.04, 1,021.81]	[-39.13, 2,403.92]
Inversion test	CLR	CLR	AR
Observations	982	982	982
Controls	LASSO	BL Risk	BL Risk

Panel C: Interacted estimates leaving 1 arm out

	(1) High treatment	(2) Low treatment	(3) Control
VSL	213.28	178.31	394.52
	(87.39)	(174.54)	(257.04)
F-statistic	82.35	23.50	16.07
Confidence Set	[43.23, 390.86]	[-170.17, 559.50]	[-88.12, 1,038.34]
Observations	963	905	982

Standard errors in parenthesis.

Panel A reports regressions using observations in all experimental arms. Panel B excludes control observations. “Interacted” columns use a first stage consisting of treatment and treatment interacted with baseline beliefs as instruments. Columns labeled “Treatment only” use only treatment assignment. Panel C reports “interacted” estimates leaving one experimental arm out, listed in the column title. All estimates control for baseline risk beliefs. Column (4) reports estimates that instrument for beliefs using exposure to a motorcycle accident victim, with robust standard errors in parenthesis. Weak instrument robust confidence sets are reported in brackets, using CLR inversion in Panel C and the indicated tests in Panels A and B. Panel C includes enumerator fixed effects and LASSO selected controls in each regression. Excludes 35 observations where surveyors reported motorcycle taxi drivers pretending to be passengers.

Table 4: Heterogeneity in the value of a statistical life

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	1(Age > median)	asinh(Wage)	1(Wage > median)	1(health > median)	1(children)	1(digit recall score > median)	1(education > median)	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
					(172.92)
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					325.00
					(123.58)
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. Standard errors in column 5 clustered by day.

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 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% 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]
F-statistic						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% 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]
F-statistic						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

Robust standard errors in parenthesis.

Columns (1) - (4) report VSL estimates obtained using equation 11. Column (5) estimates VSL using equation 12, reporting the mean. Panel A uses empirical helmet effectiveness from Liu et al. (2008). Panel B uses Ouellet and Kasantikul (2006). Column (6) reports VSL estimates using beliefs about deaths/10,000 riders to instrument for agents' beliefs about the risk reduction of a helmet. 35 observations where motorcycle taxi drivers pretended to be passengers are excluded.

Appendix

A Belief elicitation

The survey consisted of five modules. First, we collected tracking data and demographic information about respondents. Second, we obtained information about their use of motorcycle taxis, including weekly ridership volume, trip length, trip types, and reasons for using motorcycle taxis. 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. 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 were skipped for those in the pure control group.

Prior to the first survey wave, surveyors completed a one week 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 1,461 days.

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 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 belief within the range. Piloting revealed that this two step approach helped respondents answer accurately.

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, then elicited posterior beliefs. Empirical estimates were constructed by normalizing the number of recorded deaths by the NTSA the year before the study across Kenya by the total estimated number of motorcycle trips to get a per trip risk estimate. Then we used the respondent's reports of their ridership to estimate risk over the 5 year helmet lifespan.

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 a passenger that takes 6 trips per week, 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. We first present the low treatment with the Liu et al. (2008) estimate that helmets reduce one's likelihood of dying by 42% and the high treatment with the Ouellet and Kasantikul (2006) result that helmets reduce mortality risk by 70%. 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

The primary estimates of VSL report homoskedastic standard errors and weak IV robust confidence sets. This is supported by the latent utility model presented in section 3. Recall that the two-stage least squares model which identifies VSL is given by

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

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

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

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

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

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

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

where $\Delta\hat{r}_i = P_Z \Delta r_i$. Hence, by the Law of Total Variance,

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

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 other determinants of utility. The standard deviation of willingness to pay is about 41. Even if the standard deviation of VSL were as large as the mean, the contribution to ϵ_i would be just $.022 \cdot 223$ (10% of the standard deviation of WTP), 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* 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.

For robustness, I also report estimates constructed using continuously updating GMM (CUE) with heterodekastic robust standard errors in Appendix Table A4. I use GMM since 2SLS is less efficient under heteroskedasticity and to show that results are similar under different weighting of 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 apply the Frisch-Waugh-Lovell theorem to partial out the constant in the treatment only first stage and r_{i0} from the interacted, denoting 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

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

Showing individuals are weighted linearly in priors, r_{i0}. In the interacted case,

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

Thus the interacted instruments weight 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,

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

Hence, $V\hat{S}L_{2SLS}^{TO} \neq V\hat{S}L_{2SLS}^{INT}$ 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

$$
\begin{aligned}
V\hat{S}L_{2SLS}^{TO} &= \mathbb{E}\left[VSL_i \frac{r_{i0}}{\mathbb{E}[r_{i0}\zeta_i]}\right] \\
V\hat{S}L_{2SLS}^{INT} &= \mathbb{E}\left[VSL_i \frac{r_{i0}^2 \zeta_i}{\mathbb{E}[r_{i0}^2 \zeta_i^2]}\right]
\end{aligned}
$$

Therefore one typically has $V\hat{S}L_{2SLS}^{TO} \neq V\hat{S}L_{2SLS}^{INT}$ unless ζ_i and r_{i0} are independent, for instance if $\zeta_i = c$. If individuals with low beliefs round down and those with high beliefs round up, $\mathbb{E}[\zeta_i r_{i0}] > 0$. If agents always round small probabilities up or down, $\mathbb{E}[\zeta_i r_{i0}] \neq 0$.

D Ethics Appendix

This appendix describes ethical considerations taken to prevent harming respondents.

The motorcycle helmets used in the study were manufactured by Boda Plus, a local Kenyan producer. I was referred to this brand by a road safety NGO, which believed that they were high quality. I verified that the helmets adhered to the Kenyan Bureau of Standards safety regulations before the study. Independent journalists have also concluded the Boda Plus manufactures high quality helmets. For instance Weronika Strzyzynska reports in *The Guardian* article "Africa sees sharp rise in road traffic deaths as motorbike taxis boom" (2023) that "Kenyan company Boda Plus's helmets meet high safety standards but can't compete on price with low-quality imported ones."

The first result presented to respondents, that helmets reduce mortality risk by 42% as reported in Liu et al. (2008), was selected because it is often presented to consumers by the Kenyan Government, NGOs and the UN Road Safety Fund. To give one example, the FIA Foundation cites this figure in their post "Helmets testing and awareness needed to curb Kenyan motorcycle deaths, says report supported by FIA Foundation."

The second result presented to respondents is from Ouellet and Kasantikul (2006) which finds that helmeted passengers have about a 70% lower mortality rate in Thailand. Although Liu et al. (2008) is often cited, the underlying studies used in the meta analysis are primarily from high-income settings. In such cases, motorcycles drive faster and helmets are higher quality, so there may be different effectiveness versus LMICs. I selected the Ouellet and Kasantikul (2006) based on this logic, which was the most relevant LMIC that I could find. I also considered the studies from developing countries analyzed in Liu et al. (2008), but they are older or do not report mortality rates. Liu et al. (2008) did not identify any RCT evidence about helmet efficacy, possibly due to ethical concerns with a control group, which is why randomized evidence is not used.

Ouellet and Kasantikul (2006) leverages observational accident data. This study was selected because the authors also estimate that helmets are about 50% effective in Los Angeles, aligning well with the Liu et al. (2008) estimate from high-income settings. The specific estimate from Ouellet and Kasantikul (2006) presented is based on helmets that stayed on. The authors note that when helmets were ejected, the mortality rate is higher. But, as presented in the study, this typically reflects incorrect helmet use and not effectiveness. Specifically, the study reports that about 6.8% of the unhelmeted riders died, versus 1.9% of the helmet-retained riders, resulting in the estimated efficacy of about 70% presented.

Respondents were presented with one study since interpreting two requires more time, and the variation helps to understand demand for safety. This should be thought of as two different draws from the distribution of how well helmets work, not one true and one false draw. The script for both studies informed respondents that estimates were from studies outside of Kenya so may not reflect Kenyan helmet effectiveness and that other studies may produce different results. The study received IRB approval from the University of California Committee on the Protection of Human Subjects and the Amref Health Africa IRB, which is based in Kenya.

An anonymous referee expressed the thoughtful concern that the intervention may have inadvertently harmed treated respondents by reducing their willingness to pay for helmets. I appreciate the referee's careful attention to the ethical dimensions of research. However, I interpret the findings differently. The observed decline in willingness to pay followed exposure to a well-regarded study on helmet effectiveness, which corrected overly optimistic priors. From this perspective, the intervention enabled more informed decision-making, leaving participants better off. In my view, withholding evidence on helmet effectiveness in order to encourage product uptake would raise its own ethical concerns. Moreover, the sample only included those not using helmets, so it did not cause anyone that was using a helmet to stop using one.

E Appendix Figures

A1: Survey locations

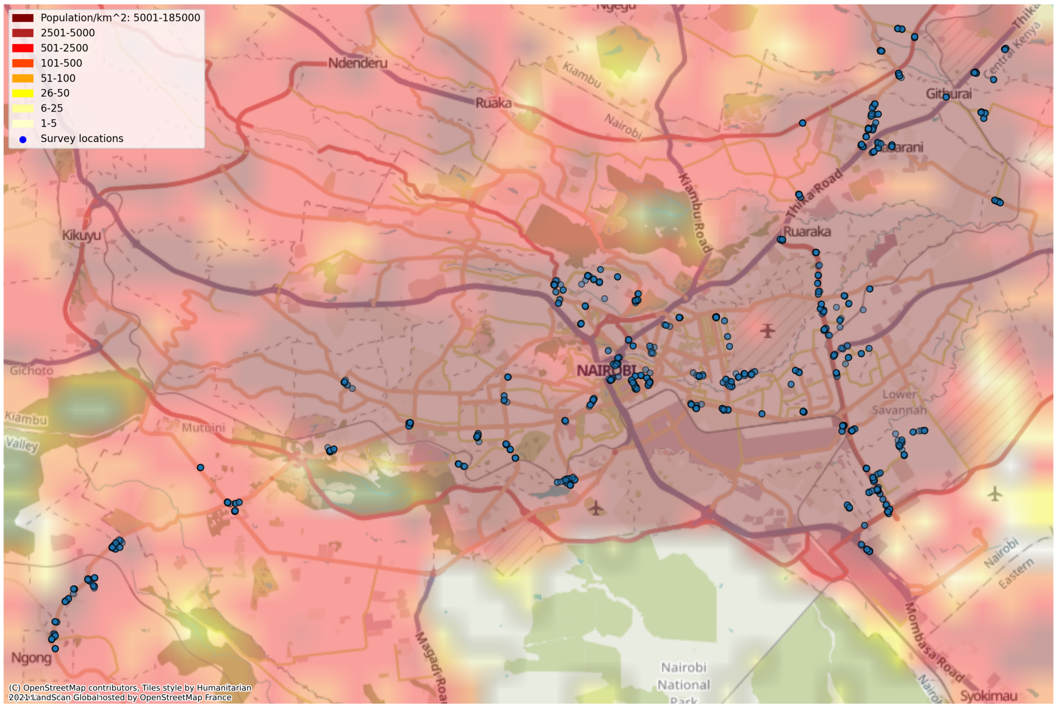
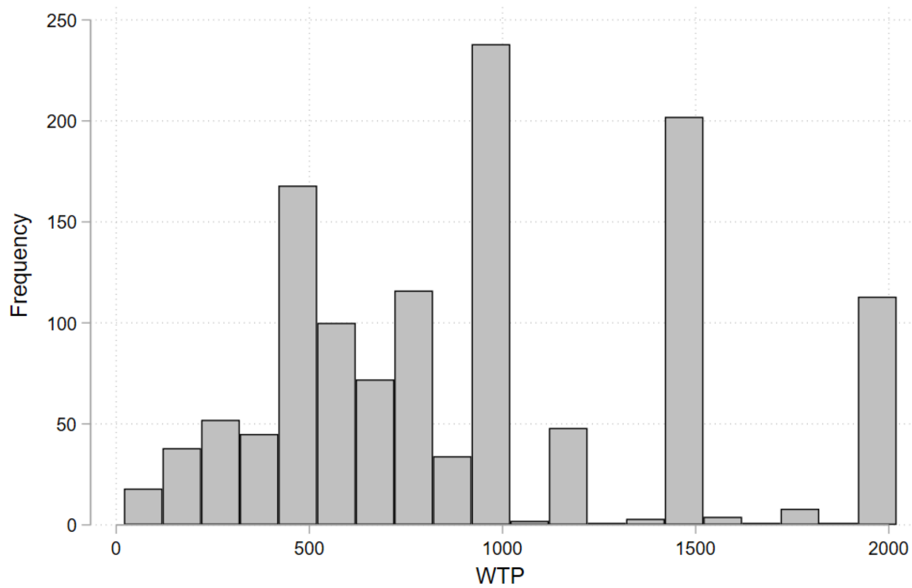


Figure A1 plots survey locations over a map of Nairobi. Color denotes population, where no color denotes no residents. The 2019 census estimated a population of about 4.4 million within the city. Map data is from Open Street Map. Population data is from the LandScan Global 2021 data set produced by Oak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

A2: Distribution of helmet bids (Kenyan shillings)

(a) A. Histogram of bids, restricted axis



(b) B. Histogram of bids, full distribution

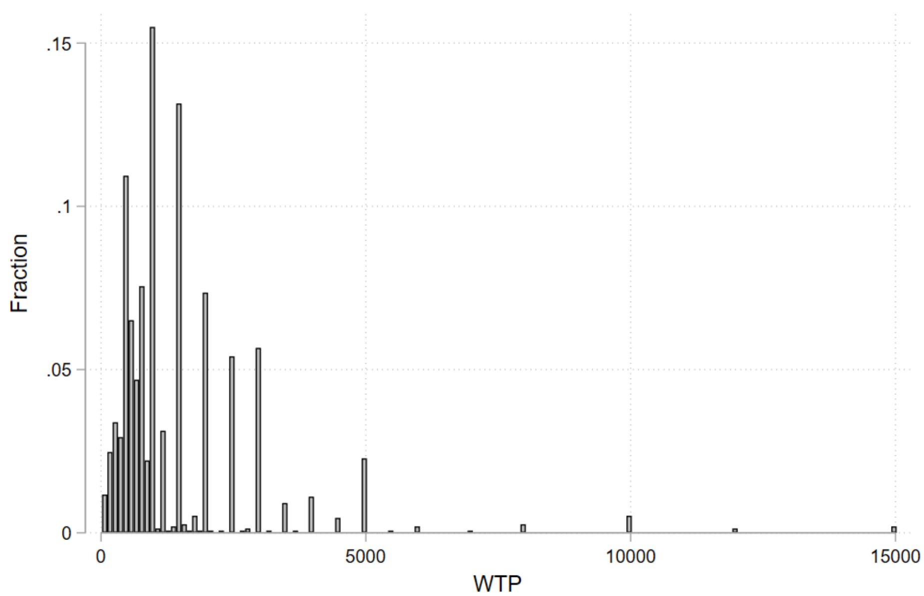


Figure A2 plots the distribution of willingness to pay for helmets in Kenyan shillings (Ksh). Panel A reports a histogram of bids, excluding outliers above Ksh 2,000 for clarity. Panel B reports the same histogram across the full distribution of bids. Figures present Ksh rather than USD to illustrate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

A3: Effect of study VSL on published benefit-cost ratios

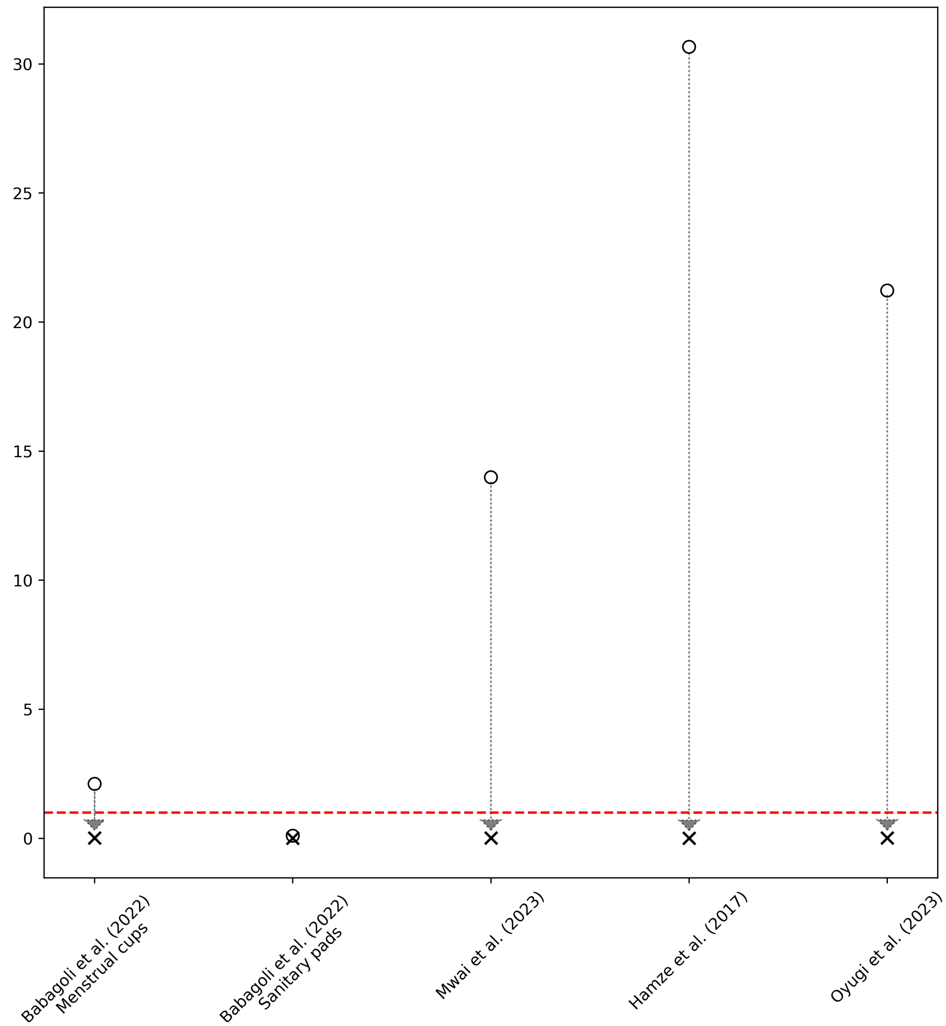
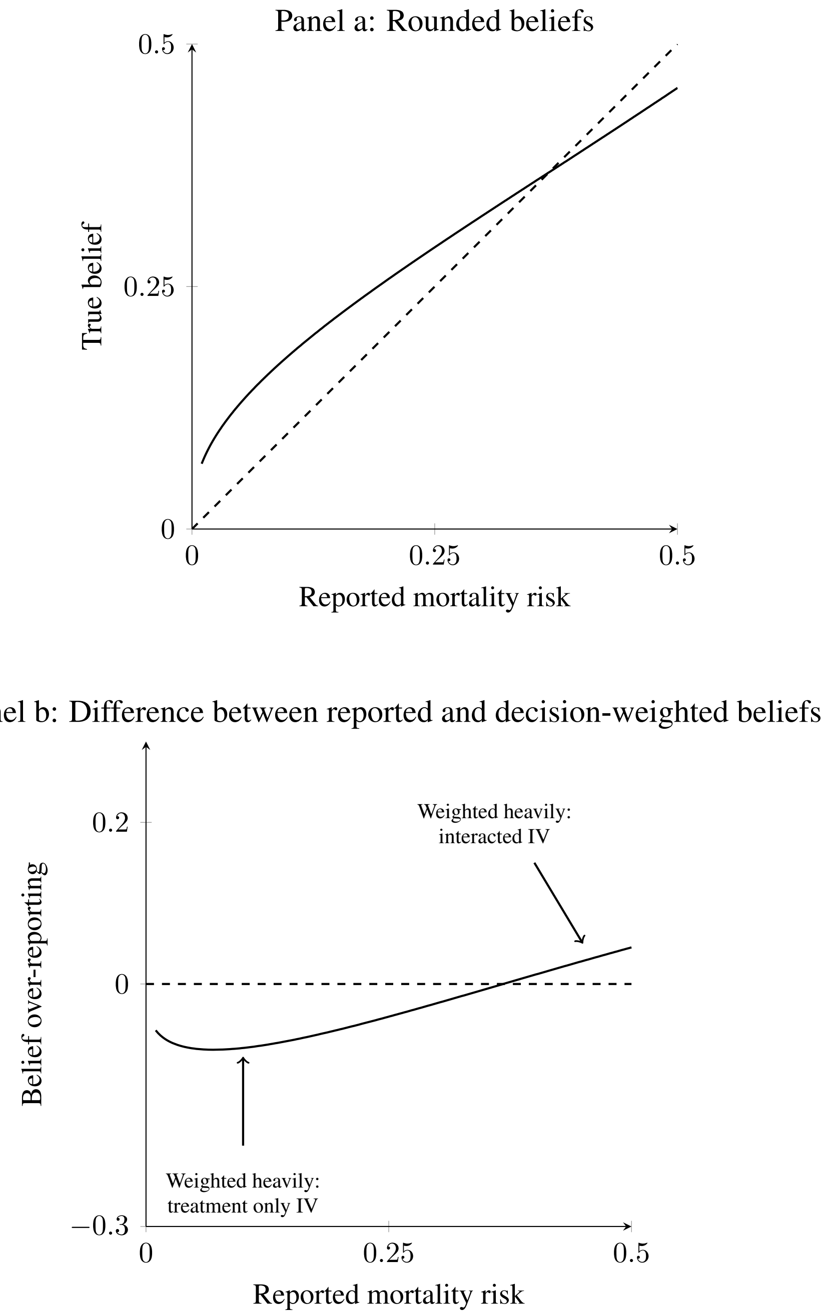


Figure A3 examines how the benefit-cost ratios (BCRs) of published benefit-cost analyses of Kenyan programs change when the study's original VSL or VSLY is replaced with the preferred estimates from this study. Hollow circles denote the original BCR estimate, and an x denotes the revised estimate. The 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.

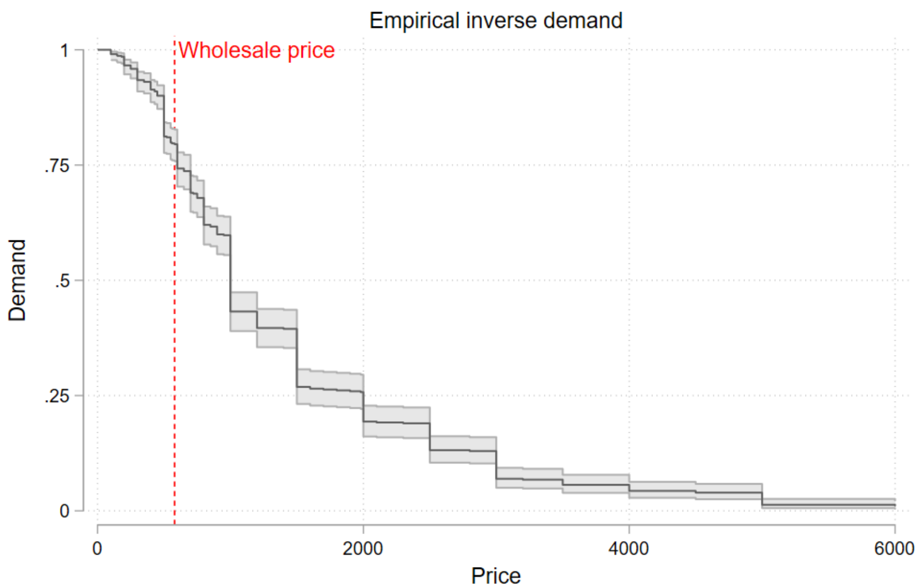
A4: Belief distortion under rounding



Appendix Figure A4 illustrates how agents with low beliefs rounding down and those with high beliefs rounding up would produce different estimates between the "treatment only" and "interacted instruments." This figure plots a stylized case where agents that report they do not believe they face any risk of dying on a motorcycle act on a belief where they do face some risk, and those that report extremely high risks act on lower values. As shown in panel b, this would induce a positive correlation between misreporting and reported risk. The "interacted" instruments would therefore weight individuals with over-reported beliefs more heavily and overestimate VSL versus the "treatment only" instruments.

A5: Demand for helmets

(a) Inverse demand, control group



(b) Elasticity of demand, control group

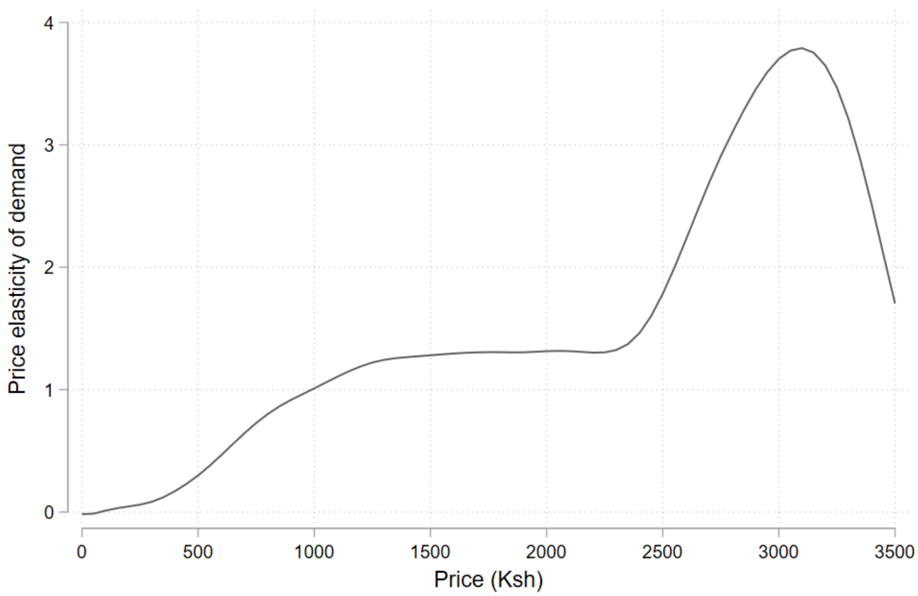


Figure A5 plots estimates of demand for helmets among observations in the control and pure control arms. The plot of demand includes a pointwise confidence interval. The plot of the elasticity of demand is based on a local polynomial estimation adapted from Berry et al. (2020). The vertical line denotes 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.

F 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: 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). 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.

A3: 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

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

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

Standard errors in parenthesis.

Panel A 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 interacted with treatment assignment. Panel B uses respondents' stated beliefs about the lifespan of the helmet, rather than the manufacturer's suggestion. Panel C weights each observation by the inverse of motorcycle trips/week to account for selection into ridership. All models control for baseline beliefs and enumerate fixed effects. Excludes 35 observations where motorcycle taxi drivers pretended to be passengers.

A4: Value of a statistical life: GMM and IV probit estimates with robust standard errors

	GMM: Raw WTP		GMM: Winsorized WTP		IV probit	
	(1) Interacted	(2) Treatment only	(3) Interacted	(4) Treatment only	(5) Treatment only	(6) Treatment only
VSL	245.74	308.12	102.70	252.59	161.25	161.92
	(141.87)	(238.51)	(76.48)	(193.40)	(118.35)	(118.42)
Price	NA	NA	NA	NA	-0.13	-0.13
					(0.03)	(0.03)
Risk reduction	NA	NA	NA	NA	20.31	20.71
					(10.17)	(10.40)
Cragg-Donald F-stat	8.16	10.96	8.16	10.96	NA	NA
Weak IV Robust Confidence Set	[-40.74, 689.51]	[-154.63, 978.63]	[-63.84, 323.75]	[-122.64, 750.35]	NA	NA
Inversion test	CLR	CLR	CLR	CLR	NA	NA
Observations	1,425	1,425	1,425	1,425	1,425	1,425
Controls	LASSO	LASSO	LASSO	LASSO	BL Risk	LASSO
Enumerator FE	Yes	Yes	Yes	Yes	Yes	Yes

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

Table A4 reports VSL estimates under heteroskedastic robust standard errors using continuous updating GMM (CUE) in columns (1)-(4) and IV probit, estimated via maximum likelihood, in columns (5)-(6). 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 (3) - (4) report estimates where willingness to pay is winsorized at the 2nd and 98th percentiles. IV probit estimates are estimated by fitting a regression of an indicator for receiving a helmet on the randomly drawn price (from the BDM game) and risk reduction, instrumented for with treatment assignment. Interacted instruments are not used for these estimates since the effective number of observations is smaller and the arguments supporting that instrument depend on linearity. Estimates exclude 35 observations for which surveyors reported motorcycle taxi drivers pretending to be passengers.