Title
Essays in Development Economics

ISBN
9798189271274

Author
Killeen, Grady Shea

Publication Date
2026-05-01

Peer reviewed|Thesis/dissertation

Essays in Development Economics

by

Grady Killeen

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Supreet Kaur, Co-chair
Professor Edward Miguel, Co-chair
Professor Benjamin Handel
Professor Jeremy Magruder

Spring 2026

Essays in Development Economics

Abstract

Essays in Development Economics

by

Grady Killeen

Doctor of Philosophy in Economics

University of California, Berkeley

Professor Supreet Kaur, Co-chair

Professor Edward Miguel, Co-chair

This dissertation comprises three essays examining how the distinctive features of developing economies shape the decisions of firms and households. The unifying argument is that economic findings from high-income settings often fail to transfer cleanly to low- and middle-income countries (LMICs), where incomplete markets, weaker infrastructure, limited social protection, and severe budget constraints fundamentally alter the problems that agents face. Ignoring these differences may harm both academic understanding of low-income economies and the effectiveness of economic development policies.

The first chapter asks whether risk aversion prevents firms in developing countries from pursuing profitable but uncertain investments. Standard economic theory assumes firms are risk neutral. This assumption may be correct in wealthy economies where owners hold diversified portfolios. In LMICs, however, most firms are owner-operated, and losses threaten household consumption directly. Using two field experiments with over 1,200 retailers in Kenya, I find that risk aversion significantly impedes the diffusion of new motorcycle helmets, harming retailers, upstream manufacturers, and consumers. These results challenge the assumption of firm risk neutrality and have broad implications for understanding innovation and growth in developing economies.

The second chapter introduces a new method for estimating the value of a statistical life (VSL) – a measure of consumers' willingness to pay for mortality risk reduction that is central to cost-benefit analyses of public policy. By experimentally updating beliefs about risk exposure without altering actual risk, I obtain a precise VSL estimate that is orders of magnitude below values from high-income settings. I argue this is theoretically consistent: as wealth declines, money's marginal value rises nonlinearly, sharply increasing the opportunity cost of investing in safety. This finding suggests that VSL estimates commonly

applied in LMIC policy contexts – typically rescaled from high-income benchmarks – may systematically misallocate resources.

The third chapter, co-authored with Michael Walker, Nick Shankar, Edward Miguel, and Dennis Egger, examines the effects of large, one-time unconditional cash transfers on infant and child mortality in Kenya. While a substantial literature documents the consumption effects of such programs, their inter-generational and health impacts remain understudied. We estimate that the transfers reduced mortality among children under five by 45%. Increases in hospital deliveries and reductions in physically demanding labor among pregnant women emerge as key mechanisms. Mortality reductions are concentrated among the poorest households, underscoring how binding budget constraints shape health outcomes at the earliest stages of life.

To my father: Timothy Francis Killeen

Contents

List of Figures

List of Tables

Acknowledgments

This dissertation would not have been possible absent the exceptional support offered my advisors: Supreet Kaur, Edward Miguel, Benjamin Handel, and Jeremy Magruder. I am grateful for the numerous hours that each of them spent providing thoughtful feedback, support, and guidance during my Ph.D.

I also owe a great deal of my success to the broader development economics and industrial organization communities at UC Berkeley. I benefited greatly from discussions with Marco Gonzalez-Navarro, B. Kelsey Jack, Gautam Rao, Frederico Finan, Carolyn Stein, Nano Barahona, and other faculty and students at Berkeley.

I would also like to acknowledge the *Georgetown University Initiative on Innovation, Development, and Evaluation* which generously implemented field research for the first two chapters of this dissertation. I am particularly grateful to William Jack, Whitney Tate, Nyambaga Muyesu, and Josephine Okello for their outstanding work in maintaining the organization and implementing field research in Kenya. The field research in this dissertation would also not have been possible without support from the Center for Effective Global Action, especially Samarth Bordia, with grant management and contracting.

I gratefully acknowledge funding from the Center for Effective Global Action, the UC Berkeley Center for African Studies, the Weiss Fund, and the Clausen Center for International Business and Policy. This document is an output from the research initiative 'Private Enterprise Development in Low-Income Countries' (PEDL), a programme funded jointly by the by the Centre for Economic Policy Research (CEPR) and the Foreign, Commonwealth and Development Office (FCDO), contract reference PEDL_LOA_10670_Kaur. The views expressed are not necessarily those of CEPR or FCDO. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 2146752.

The third chapter of my dissertation was supported by grants from the National Science Foundation, International Growth Centre, CEPR/Private Enterprise Development in Low-Income Countries (PEDL), Weiss Family Foundation, an anonymous donor, and Open Philanthropy (recommended by GiveWell). We thank Jasmin Baier, Ilaria Dal Barco, Daniel Han, Layna Lowe, Prince Muraguri, Rachel Pizatella-Haswell, and Maya Shen for excellent research assistance, and Carol Nekesa, Andrew Wabwire and REMIT Kenya for data collection and support.

Chapter 1

Risk Aversion and Barriers to Firm Growth

1.1 Introduction

Firms in low and middle-income (LMIC) economies often grow slowly and have low productivity (Bloom, Mahajan, et al., 2010). Most of the existing literature on firms, even small and poor, assumes risk neutrality, meaning only expected returns should drive decisions. Yet most production in developing economies occurs among small enterprises where owners bear a significant share of profits and losses. In the absence of complete insurance markets, owners' consumption may be sensitive to business performance, causing their risk preferences to affect production decisions. This may deter enterprises from engaging in risk taking required to grow.

This paper investigates whether small firms are risk averse, causing them to forego profitable but risky investments. The returns to new business practices, such as employing a new technology or stocking a new product, are inherently uncertain. If firms are risk averse, they may fail to engage in experimentation to identify profitable opportunities, limiting innovation and growth.

I explore this question in the context of retail firms' decision to stock a new consumer product. Offering new products can increase retailers' profits. But shops are often uncertain about whether consumers will demand new goods, creating a risk that inventory investments will lead to losses. This setting also has direct implications for upstream producers. A manufacturer who cannot secure retail distribution faces a smaller effective market, reducing the returns to introducing new goods. If risk aversion deters retailers from stocking new products, it may therefore act as a barrier not only to firm growth but also to the introduction of new products in developing economies.

My empirical setting is the Kenyan market for new motorcycle helmets. In 2020, a corporation built a USD $3 million Kenyan factory and began producing one of the first effective helmets affordable to consumers, selling at a wholesale price of USD PPP 15 (high-quality imports cost $100 or more). But their retail diffusion was slow. Two years after their introduction, over 50% of shops near the factory believed that selling helmets would yield positive expected profits, yet only 6% stocked them. Although many firms held optimistic expectations, their beliefs were often diffuse, and retailers did not stock the good if they were unsure about its profitability.

I begin by developing a model that draws stark contrasts between the experimentation and learning behavior of risk neutral vs. risk averse firms when faced with an uncertain investment. In the model, an entrepreneur decides whether to stock a new good with unknown demand. This embeds a bandit learning problem into a standard model of small firm behavior, generating predictions identifying risk aversion and demand uncertainty. The core insight is that when firms are more uncertain about a product's profitability, risk neutral firms invest more because upside is high. This ensures that risk neutral retailers quickly discover profitable new products. In contrast, risk aversion creates disutility from uncertainty, which can lead firms to invest less in products the more uncertain they are about their profitability, undermining new product adoption.

Guided by the model, I design two field experiments to test if risk aversion prevents enterprises from efficiently adopting new products. The first experiment, the Insurance Experiment, constructs a precise test for whether risk aversion affects firms' stocking decisions. The second experiment, the Learning Experiment, isolates the importance of risky experimentation to learn in an uncertain environment. This experiment shows that temporarily lowering the risk of experimenting leads to permanent increases in helmet stocking. I leverage a treatment arm and model predictions to argue that persistence is driven by risk averse firms overcoming uncertainty and cannot be explained by confounding explanations such as incorrect beliefs or learning by doing.

The Insurance Experiment ($N = 350$ firms) tests if risk aversion affects firms' stocking decisions. I offered shops the opportunity to stock helmets with or without an insurance contract designed to be strictly dominated under risk neutrality but valuable under risk aversion. I test if enterprises are risk neutral by examining the effect of this contract on stocking.

Testing firms' risk preferences is complicated by possible capital constraints, since insurance contracts typically charge upfront premiums, affecting liquidity. To isolate risk aversion, my design offered control shops an unconditional future payment. Randomly selected *insurance* shops chose between this payment and a contract paying more if they stocked helmets and failed to sell out, but nothing otherwise. Offers were calibrated so that, given firms' beliefs, the contract's expected value fell below the fixed payment, creating a mean-preserving (technically decreasing) contraction of helmet profits. Therefore, *insurance* should increase stocking only if firms are risk averse.

The rate of *insurance* shops that immediately stocked helmets was 7 percentage points (157%, $p = .01$) higher. Firms were given up to two weeks to order to search for customers. The *insurance* effect was 10pp (50%, $p = .02$) after this period. This implies a rejection of risk neutrality. The results are not driven only by new or unproductive firms: effects remained large and significant among firms in business for at least 4 years, with employees, and with above median profits.

The design is robust to two additional confounds. First, surveyors followed a script to inform all shops about *insurance* and to explain that the shops receiving the offer would be selected at random. Then the shop's treatment assignment was revealed. This prevents *insurance* from sending a signal about demand. Second, the contract included conditions ensuring that stocking helmets and intentionally failing to sell out to obtain the *insurance* payout was never optimal.

The Learning Experiment ($N = 929$ firms) tests if risk aversion prevents firms from engaging in experimentation needed to discover profitable new products. The design builds on the model's prediction that firms face a risk-reward trade-off: by stocking a new product, they risk short-term losses but may gain long-term profits by restocking the product if they learn that it is profitable. The sample consists of firms that could have acquired helmets from the manufacturer for over two years but did not stock them. Two treatments test whether risk aversion prevented firms from learning about helmets' profitability in this setting where their background diffusion was slow.

The first *returns* treatment was designed to test if temporarily lowering the risk of selling helmets caused firms to learn about demand and permanently stock them. All firms received access to helmet stock at market prices. The intervention had two phases. In phase one, *returns* shops could return unsold stock for a refund whereas control shops could not. This mirrors common policies in high-income settings (Li and Kim, 2022). In phase two, all firms were given the option to buy new stock without a return option, so that offers were identical for *returns* and control shops. The key test is whether the one-time reduction in risk led to increases in phase two stock purchases.

Consistent with the Insurance Experiment, the temporary reduction in inventory risk led to a large increase in helmet uptake in phase one: 16.4% of *returns* shops stocked helmets compared to 6.8% of control firms (241%, $p < .01$). Reflecting firms' baseline expectations that the product would be profitable, few firms exercised the return option and 90% of shops directly reported that they were profitable, with magnitudes amounting to 10% of total firm profits.

These initial uptake differences translated into persistently higher rates of helmet stocking in phase two, when all firms faced the same stocking conditions. *Returns* shops were 7 percentage points (70%, $p = .03$) more likely to purchase helmet stock (from any supplier) and twice as likely to purchase inventory at least twice. This was driven by high restocking rates consistent with reported profitability: 2/3 of phase one adopters restocked and 80% planned to stay in the market. The increase in phase two stocking is consistent with *returns* inducing risk averse firms to try selling helmets, who then resolved demand uncertainty by learning.

A confounding explanation is that shops held incorrect beliefs that helmets were unprofitable, and the intervention positively updated their expectations. The model predicts divergent patterns in firms' beliefs under these two mechanisms, which I test using novel firm survey data.

If returns mitigate risk aversion, (1) belief variance should negatively predict uptake among untreated firms (stocking is perceived as risky), (2) *returns* should induce stocking among firms uncertain about demand, and (3) beliefs should become more precise with experience. Under the alternative of changing mean beliefs, (1) uncertainty should positively predict uptake among untreated firms (information is more valuable), (2) the treatment should attract firms with pessimistic priors, and (3) beliefs should become more optimistic with experience. The empirical evidence supports the importance of risk aversion: all three predictions associated with this mechanism hold, versus none under the alternative of positively updating expectations.

These results highlight the importance of firm learning under uncertainty. When experimentation is risky but can generate long-run gains, a firm's willingness to experiment depends not only on risk but also on the perceived value of future opportunities. This implies an additional prediction of the model: firms should be more willing to undertake experimentation in the present when they anticipate larger payoff potential in the future, even if current-period payoffs are unchanged.

The Learning Experiment tests this hypothesis using a second *supplier commitment* treatment. Firms expressed concerns that the manufacturer may stop supplying small shops in the future, reducing future upside from discovering if helmets were profitable. I therefore offered, during phase one, to help *supplier commitment* shops restock directly from the supplier after the study. As in the Insurance Experiment, control shops were aware of this arm but informed that they would need to identify a supplier themselves. *Supplier commitment* firms were 6 percentage points (84%, $p = .03$) more likely to stock helmets during phase one, matching with the model's prediction.

This finding shows that firms are sophisticated and forward-looking. This suggests that low product experimentation absent *returns* does not simply reflect firms failing to internalize the learning value of stocking. Rather, the results are consistent with risk aversion creating high utility costs that deter experimentation, constraining firm innovation and productivity growth.

The model further predicts that *returns* and the *supplier commitment* should be substitutes, helping rule out mechanisms that could confound this interpretation. If *returns* effectively eliminate risk, then all firms that believe helmets may be profitable should stock. Therefore uptake should not be affected by also receiving the *supplier commitment*. A subset of firms received both offers to test this prediction. Results align with the model: the *supplier commitment* has no effect conditional on receiving *returns*. This helps to rule out confounding forces, such as learning-by-doing, that would typically cause the *supplier commitment* to affect uptake even with *returns*.

Another implication of demand uncertainty as the key source of risk constraining helmet adoption is that information acquired from other sources should also affect stocking. I document evidence of information spillovers across firms consistent with this view. These externalities also help to explain why competition does not drive risk averse firms out of business, suggesting that it may be possible for shops to free ride off the product experimentation of less risk averse competitors.

Three results highlight spillovers. First, *returns* had no effect when firms were located near an existing seller. In such cases, uptake was common in all arms, consistent with learning from incumbent sellers. Second, providing a random group of firms with peers' helmet sales data increased stocking by 2 percentage points ($p = .02$). Third, I induced helmet sellers to enter selected treated markets and later offered a new *spillover* sample helmet stock in treated and control markets. Firms in treated markets were nearly twice as likely to purchase helmet stock ($p < .01$).

Finally, the expansion of the helmet market produced by the Learning Experiment was economically important, suggesting that risk aversion can consequentially constrain the retail market for new products and reduce the revenues earned by upstream producers. *Returns* shops stocked twice the number of helmets during the study (3.4 vs 1.5, $p = .03$), and five times the volume in markets with no pre-existing seller (3.5 vs 0.7, $p < .01$). Moreover, *returns* enterprises remained 8 percentage points ($p = .04$) more likely to sell helmets or report a nearby vendor at endline. The magnitude of these effects two years after the helmet factory was built is reconciled with externalities by the fact that spillovers are highly localized: survey data shows shops rarely observed helmet stocking among firms more than 0.25 km from them (about 3 blocks), limiting information diffusion.

This study's principal contribution is to the literature studying barriers to firm growth in developing countries. Prior work has studied how inputs such as capital (Mel, McKenzie, and Woodruff, 2008), labor (Mel, McKenzie, and Woodruff, 2019), and management (Bloom, Eifert, et al., 2013) affect productivity. This paper provides the first direct evidence that risk aversion can prevent firms from engaging in entrepreneurial risk taking that is necessary to grow, lowering uptake of a profitable product. I also introduce a new test of risk aversion that addresses key confounding factors and designed survey data to distinguish from competing models. This builds on prior studies examining correlations between risk preferences and firm outcomes (Mel, McKenzie, and Woodruff, 2008; Kremer, Lee, et al., 2013; Meki, 2025), evidence that farmers are risk averse (Karlan et al., 2014), and observations that risk preferences may rationalize features of small enterprise behavior (Pelnik, 2024).

This finding implies that the common practice of modeling small firms as risk neutral may not be appropriate. Risk neutrality plays an important role in constructing empirical tests and interpreting findings in a vast range of topics within development economics, including capital constraints, collusion, technology adoption, and misallocation (Fafchamps et al., 2014; Bergquist and Dinerstein, 2020; Bassi et al., 2022; Buera, Kaboski, and Shin, 2011). However, if risk aversion means that firm behavior cannot be interpreted solely through the lens of profit maximization, the interpretation of tests that are standard across these varied topics may change. For instance, the marginal product of capital may vary across firms due to heterogeneous risk preferences, not only borrowing access.

The results also show that risk aversion and demand uncertainty can slow the diffusion of new goods. This is important to economic development, as the slow adoption of products and technologies is widely recognized as an important constraint to economic growth (Comin and Mestieri, 2018). Slow retail diffusion may also reduce manufacturers' incentives to introduce new products in developing markets. These findings build on research on barriers to technology diffusion and firm upgrading in developing economies (Atkin et al., 2017; Verhoogen, 2021), extending results to retail settings. More broadly, the finding that a temporary return intervention permanently boosts product adoption suggests that such policies could be a cost effective way to promote diffusion.

A further contribution of this study is to a literature on learning about demand. The model of small firm learning developed in this paper expands on a theoretical literature examining demand uncertainty (Rothschild, 1974; Bolton and Harris, 1999) and on empirical studies from other contexts (Foster and Rosenzweig, 1995; Doraszelski, Lewis, and Pakes, 2018). The model highlights that risk neutral firms should hold accurate beliefs about demand in equilibrium because uncertainty incentivizes experimentation with goods. However, I show that risk aversion can prevent shops from experimenting, rationalizing evidence that firms in developing countries can hold inaccurate beliefs about the profitability of products (Bai, Park, and Shenoy, 2025). I also show evidence that firms internalize the learning value of investments, empirically validating a key prediction of bandit models of learning.

1.2 A model of small firm learning about demand

In this section, I model an entrepreneur faced with the decision of whether or not to stock a new product. The model embeds the problem of learning about demand for a new product into the optimization problem of a small firm owner. I begin by describing the problem faced by the agent, then derive equilibrium conditions. The section concludes by constructing tests of risk aversion and learning about demand that guide the design and interpretation of the experiments.

Model setup

I consider an infinitely repeated, single agent dynamic optimization problem in discrete time. The model is single agent because the hypotheses studied are not strategic. The problem is dynamic, as learning enables agents to refine their future decisions.

The entrepreneur's problem: The entrepreneur chooses how much of a safe product $j = s$ and a new product $j = n$ to stock in each period t. Inventory is sold in period $t + 1$. The agent knows the residual demand curve of the safe good $p_s(q_{st}, \nu_{st})$, but realized demand is subject to iid stochastic fluctuations $\nu_{st} \sim \mathcal{N}(0, \Sigma_s)$.

Demand for the new product, $p_n(q_{nt}, \nu_{nt} + \theta)$, is indexed by a parameter $\theta \in \mathbb{R}^k$. The true value, θ_0, is unknown to the agent. $\nu_{nt} \sim \mathcal{N}(0, \Sigma_n)$ again captures iid fluctuations in demand. These influence the rate at which agents learn and allow for a distinction between two sources of uncertainty: diffuse prior beliefs, which can be resolved through learning, and stochastic demand fluctuations, which introduce variance in profits even when the demand curve is fully known.

The demand curves are continuously differentiable, downward sloping $\left(\frac{\partial p_j}{\partial q_{jt}} \leq 0 \ \forall q_{jt}, \nu_{jt}\right)$ and continuously differentiable and increasing in ν_{jt}. p_{jt} and q_{jt} are observed, but ν_{jt} is not observable.

Flow profits from the safe good, conditional on stocking q_{st}, are given by $\pi_{st} = q_{st}p_j(q_{st}, \nu_{st}) - \zeta_s(q_{st})$ where $\zeta_s(\cdot)$ is a known and differentiable function capturing non-stock costs of selling q_{st}, such as labor and capital used for marketing. Flow profits from the new product are determined by $\pi_{nt} = q_{nt}p_j(q_{nt}, \nu_{nt}; \theta) - \zeta_n(q_{nt})$ where $\zeta_n(\cdot)$ is also known and differentiable. Agents may invest in any non-negative stock of the safe product each period, $I_{st} \geq 0$ at wholesale cost w_s per unit. Wholesalers impose a minimum order size χ for the new product, so $I_{nt} \in \{0\} \cup [\chi, \infty)$.[1] The new product has a wholesale price of w_n. Since inventory is sold each period, $q_{jt} = I_{jt-1}$.

In addition to demand uncertainty, agents face supply chain uncertainty when stocking the new product. This captures concerns that unfamiliar manufacturers may fail or shift focus, cutting small firms off from future supply. To capture this, the model includes a fixed continuation cost Γ that the agent expects to incur in period $t_c > 1$ if they wish to keep stocking the new product. This provides a way to study how agents respond to changes in the continuation value of learning.

Learning about demand: A core feature of the model is that the agent learns about new product demand. They begin with a prior $\theta \sim \mathcal{N}(\mu_1, \Sigma_1)$ and Bayesian update when their information set, $\mathcal{I}_t(I_{n0}, \dots, I_{nt-1})$, changes. If the agent stocks I_{nt}, then in period $t+1$ they receive a signal $x(I_{nt}) \sim \mathcal{N}\left(\theta_0, I_{nt}^{-1}\Sigma_x + \Sigma_n\right)$. The agent knows that the signal is centered around the truth and knows its precision, but does not know θ_0. The precision of information about demand the agent receives is increasing in their level of investment, reflecting the fact that a greater stock provides more opportunities for customer interactions, price experimentation, and data about sales. However, the learning becomes more difficult if demand fluctuates substantially from period to period.

The information set depends on time t because agents may also learn from external sources, such as neighboring retailers. Each period the retailer receives a signal $x_{ot} \sim \mathcal{N}(\theta_0, \Sigma_o)$ with probability φ where Σ_o (o for "other source") is known. Beliefs update according to Bayes' Rule.

$$\begin{aligned}
&\theta_t \sim \mathcal{N}(\mu_t, \Sigma_t) \\
&\mu_t = \Sigma_t \left(\Sigma_1^{-1}\mu_1 + \sum_{\tau=1}^{t-1} (I_{n\tau}^{-1}\Sigma_x + \Sigma_n)^{-1} x(I_{n\tau}) + \sum_{x_{ot} \in \mathcal{I}_t} \Sigma_o^{-1} x_{ot} \right) \\
&\Sigma_t = \left(\Sigma_1^{-1} + \sum_{\tau=1}^{t-1} (I_{n\tau}^{-1}\Sigma_x + \Sigma_n)^{-1} + |x_{ot} \in \mathcal{I}_t| \Sigma_o^{-1} \right)^{-1}
\end{aligned} \tag{1.1}$$

The agent's objective: The entrepreneur receives flow utility of consumption given by a continuously differentiable function $u(\cdot)$. They may save or borrow at interest rate r and

[1] This condition captures realistic features of the market and allows for an equilibrium in which the agent does not fully learn demand. Imposing fixed delivery costs without a minimum order size yields similar results.

discount the future at rate $\delta = \frac{1}{1+r}$ and are subject to borrowing limit $\underline{a} \leq 0$, capturing possible capital constraints. The agent begins with assets $a_0 > 0$ but no stock. Their objective is

$$\max_{\{c_t, a_t, I_{st}, I_{nt}\}} \mathbb{E}_0 \sum_{t=1}^{\infty} \delta^{t-1} u(c_t) \tag{1.2}$$

subject to a budget constraint $a_t + c_t + w_s I_{st} + w_n I_{nt} \leq (1+r)a_{t-1} + \pi_s(I_{st-1}, \nu_{st}) + \pi_n(I_{nt-1}, \nu_{nt} + \theta_0)$, minimum order sizes of the new product $I_{nt} = 0$ or $I_{nt} \geq \chi$, the borrowing limit $a_t \geq \underline{a}$, non-negative investment $I_{st} \geq 0$, and a transversality condition. Expectations are over θ, due to incomplete information, and ν_{jt}, due to stochastic variation in demand.

The value of learning affects an agent's new product stocking because investment facilitates learning, which allows better future optimization. This enters their objective as reductions in future "regret," lost utility due to incomplete knowledge of θ. Let $y_t = (1+r)a_{t-1} + \pi_s(I_{st-1}, \nu_{st}) + \pi_n(I_{nt-1}, \nu_{nt} + \theta_0)$ be the agent's cash on hand. Define the conditional value function

$$V^*(y_t, \Gamma, \theta) = \max_{\{c_\tau, a_\tau, I_{s\tau}, I_{n\tau}\}} \sum_{\tau=t+1}^{\infty} \delta^\tau \, \mathbb{E}[u(c_{t+\tau})|\theta] \tag{1.3}$$

subject to the some conditions as Equation 1.2, but treating θ as known. Let $\bar{c}_t$ denote consumption along the path that solves the original objective, maximizing over beliefs about θ instead of treating it as known. Define

$$V(y_t, \mathcal{I}_t, \Gamma, \theta) = \sum_{\tau=1}^{\infty} \delta^\tau \, \mathbb{E}[u(\bar{c}_{t+\tau})|\theta] \tag{1.4}$$

This is the expected utility the agent will receive from their planned actions if $\theta_0 = \theta$.

Regret is given by $R(y_t, \mathcal{I}_t, \Gamma, \theta) \equiv V^*(y_t, \theta, \Gamma) - V(y_t, \mathcal{I}_t, \theta, \Gamma) \geq 0$. This captures lost utility due to incomplete information about demand if $\theta_0 = \theta$. An agent's Bayesian regret is

$$\bar{R}(y_t, \mathcal{I}_t, \Gamma) \equiv \mathbb{E}_\theta \left[R(y_t, \mathcal{I}_t, \Gamma, \theta) | \mathcal{I}_t \right] \tag{1.5}$$

Which is the lifetime utility that the agent expects to lose because of uncertainty about θ.

Model solution

The solution to the model is derived in appendix A.1. I present and interpret the results for optimal investment in this section.

Optimal investment in the safe good: The utility maximizing level of investment in the safe good, I^*_{st}, satisfies

$$\begin{aligned} &\delta \left\{ \mathbb{E}_t \left[\frac{u'(c_{t+1})}{u'(c_t)} \right] \mathbb{E}_t \left[\frac{\partial}{\partial I_{st}} \pi_s(I^*_{st}, \nu_{st+1}) \right] \right. \\ &\left. + \frac{1}{u'(c_t)} Cov_t \left(u'(c_{t+1}), \frac{\partial \pi_s(I^*_{st}, \nu_{st+1})}{\partial I_{st}} \right) \right\} = w_s - \frac{1}{u'(c_t)} \iota_{st} \end{aligned} \tag{1.6}$$

where $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot|\mathcal{I}_t]$ and $Cov_t(\cdot) = Cov(\cdot|\mathcal{I}_t)$. $\mathbb{E}_t\left[\frac{u'(c_{t+1})}{u'(c_t)}\right]$ will equal 1 when capital constraints do not bind and captures the fact that investment will fall when agents hit their borrowing limit, in which case $u'(c_t)$ will be greater than $\mathbb{E}_t[u'(c_{t+1})]$ so the shadow cost of investment increases. ι_{st} is a Lagrangian multiplier ensuring non-negative investment.

$Cov_t\left(u'(c_{t+1}), \frac{\partial \pi_s(I^*_{st}, \nu_{st+1})}{\partial I_{st}}\right)$ captures possible risk aversion and will be zero if agents are risk neutral ($u'(\cdot)$ is then constant). Low profits reduce consumption (if insurance markets are incomplete). If $u(\cdot)$ is concave, this covariance will be negative and increasing in magnitude with the variance of profits and the agent's risk aversion, leading to inefficiently low stocking levels.

This argument relies in part on the fact that the model features credit constraints, which induces convexity of the marginal utility of consumption. If agents had sufficient buffer stock savings such that negative shocks could never cause them to hit their borrowing limit, then under quadratic utility certainty equivalence would hold and prudence ($u'''(\cdot) > 0$) would be needed to induce a distortion in stocking. Because agents are often credit constrained in the empirical setting, this paper does not attempt to disentangle prudence from buffer stock savings concerns.

Optimal investment in the new good: I^*_{nt} is defined by the Euler equation

$$\delta\Bigg\{\underbrace{\mathbb{E}_t\left[\frac{u'(c_{t+1})}{u'(c_t)}\right]}_{\text{Capital constraints}}\underbrace{\mathbb{E}_t\left[\frac{\partial}{\partial I_{nt}}\pi_n(I^*_{nt}, \nu_{nt+1}+\theta)\right]}_{\text{Expected marginal profits}} + \underbrace{\frac{1}{u'(c_t)}Cov_t\left(u'(c_{t+1}), \frac{\partial \pi_n(I^*_{nt}, \nu_{nt+1}+\theta)}{\partial I_{nt}}\right)}_{\text{Risk aversion}}$$
$$-\underbrace{\frac{1}{u'(c_t)}\delta\,\mathbb{E}_t\left[\frac{\partial \bar{R}(y_{t+1}, \mathcal{I}_{t+1}, \Gamma)}{\partial I^*_{nt}}\right]}_{\text{Marginal learning value}}\Bigg\} = \underbrace{w_n - \frac{1}{u'(c_t)}\kappa_{\chi t}(2I_{nt}-\chi)}_{\text{Marginal investment costs}} \tag{1.7}$$

given that the utility of I^*_{nt} satisfies the participation constraint that it exceeds that of not stocking n. Since a change in primitives which increases I^*_{nt} also relaxes the participation constraint, I focus on Equation 1.7 to study extensive and intensive margin decisions.

There are two important differences from equation 1.6. First, expected profits and the covariance between profits and marginal utility depend on θ, and beliefs about this parameter change with information. For instance, if the agent receives a signal centered on their prior, expected profits will be unchanged but a risk averse agent will perceive less stocking risk, reducing this term in magnitude. Thus learning can overcome demand uncertainty.

Second, learning has value because it reduces regret. Let $V_{nt} = I^{-1}_{nt}\Sigma_x + \Sigma_n$ be the variance of the signal the agent receives. Letting $R_\tau(\theta) = u(c^*_\tau(\theta)) - u(\bar{c}_\tau(\theta))$, Appendix A.1

shows

$$\mathbb{E}_t\left[\frac{\partial \bar{R}(y_{t+1}, \mathcal{I}_{t+1}, \Gamma)}{\partial I^*_{nt}}\right] = -\frac{1}{2}\sum_{\tau=t+2}^{\infty}\delta^{\tau-2}Cov_t\left(\underbrace{R_\tau(\theta)}_{\substack{\text{Sensitivity of}\\ \text{utility to }\theta}}, \underbrace{(\theta-\mu_\tau)' I_{nt}^{-2} V_{nt}^{-1}\Sigma_x V_{nt}^{-1}(\theta-\mu_\tau)}_{\text{Reduction in uncertainty from investment}}\right) \leq 0 \tag{1.8}$$

In words, the increase in future utility one obtains from stocking more of the new product is a function of how strongly utility varies with θ and the sensitivity of learning to investment. Agents will stock past the point where the immediate payoff is 0 because investment allows them to better optimize in the future. This is the "risk-reward" trade-off typical of bandit models.

For instance, consider a risk neutral agent that expects new product profits to be barely negative, is uncertain about demand, and can immediately learn demand by stocking. Then $I^*_{nt} > 0$ because the firm can exit the market if the product is not profitable but restock if it is. The present value of possible restocking profits will exceed the one period expected losses, the cost of experimenting.

A core insight of this model is that the value of learning is increasing in uncertainty about demand, $|\Sigma_t|$, because high uncertainty implies a probability that the product is very profitable. As a result, a risk neutral firm's willingness to stock a new product is increasing in demand uncertainty, ensuring that profitable goods are discovered. However, for risk averse firms, the potential downside associated with uncertain payoffs creates disutility, constraining experimentation. If risk aversion is sufficiently high, this relationship may reverse: greater uncertainty can prevent firms from investing. Thus, deviations from risk neutrality can fundamentally impede product diffusion.

Two additional insights inform the predictions used to test the model. First, supply chain uncertainty, Γ, lowers the continuation value of learning without affecting short-run returns, so it may be used to test the "risk-reward" trade-off. Beginning in period t_c, the agent would only stock $I^*_{nt} > 0$ if the present value of expected lifetime utility gains from stocking it exceeded Γ. An increase in Γ thus reduces $\mathbb{E}_t\left[\frac{\partial \bar{R}(y_{t+1}, \mathcal{I}_{t+1}, \Gamma)}{\partial I^*_{nt}}\right]$ in magnitude. But profits are unaffected for $t < t_c$.

Second, indirect tests of risk aversion used in past work – such as examining heterogeneous returns to capital or portfolio choices – are sensitive to modeling assumptions. For instance, Mel, McKenzie, and Woodruff (2008) model an enterprise stocking a single good with uncertain demand that cannot be overcome by learning. Given this structure, risk averse firms are further from the efficient stocking level and have higher returns to capital. However, this test is ambiguous when firms stock multiple products: more risk averse agents may invest funds in safer products with lower expected returns. This paper instead focuses on predictions relating to the response of new product stocking to changes in risk, which yields robust tests of risk neutrality.

Comparative static predictions

I next derive testable predictions of the model. I first derive three propositions that examine whether risk aversion and demand uncertainty consequentially affect new product adoption. A fourth proposition then considers information externalities.[2]

Definition 1 (Mean-preserving contraction) *A mean-preserving contraction of the profit function at the minimum order size, χ, is a one-period perturbation, $\pi_n^p(\cdot)$, such that*

1. $\mathbb{E}_t\left[\pi_n^p(\chi, \nu_{nt+1}+\theta)\right] = \mathbb{E}_t\left[\pi_n(\chi, \nu_{nt+1}+\theta)\right]$
2. $\int_{-\infty}^{x} F_p(\chi, y)dy \leq \int_{-\infty}^{x} F(\chi, y)dy \quad \forall x$*, and strictly for some x. $F(I_{nt}, y)$ denotes the probability that $\pi_n(I_{nt}, \nu_{nt+1}+\theta) \leq y$, and F_p is the same for the perturbed profit function.*
3. $\pi_n^p(I_{nt}, \nu_{nt+1}+\theta) = \pi_n(I_{nt}, \nu_{nt+1}+\theta) \quad \forall I_{nt} \neq \chi$

Expected profits are unaffected by a mean-preserving contraction, but the distribution is less spread out.[3] Since expected profits are unchanged, this will not affect a risk neutral firm's decision. However, a risk averse agent's expected utility of stocking will increase.

Proposition 1 (Risk aversion) *A mean-preserving contraction leads a firm to stock $I_{nt}^* > 0$ that otherwise would not if and only if they are risk averse.*

This provides an unambiguous test for firm risk aversion that is not affected by capital constraints. This result is sensitive to the fact that the contraction lasts for one period, which ensures that it does not reduce the continuation value of learning.

I next derive tests of learning about new product demand by analyzing the model's predictions about two policies: (i) a one-time return offer, and (ii) a reduction in supply chain uncertainty. A return policy guarantees that the firm receives at least the good's stocking cost ($p_{nt+1} \geq w_n$), reducing downside risk and increasing the expected marginal return from stocking under uncertainty. A reduction in supply chain uncertainty, Γ, increases the continuation value of learning. Therefore both policies should increase new product investment if demand uncertainty binds.

A sharper implication arises from the interaction between the return offer and the reduction in Γ, which helps distinguish targeted responses from alternative explanations such as learning-by-doing, fixed costs, or other omitted dynamics. If the return policy fully insures the firm against losses, then any agent who believes the product could be profitable will stock it. Under this condition, a reduction in Γ has no additional effect (at the extensive margin) if it operates through the targeted channel. In contrast, if supply chain frictions

[2] I focus on intuition for the propositions and their implications in the text. Proofs are presented in Appendix A.1.

[3] I focus on a mean-preserving contraction at the minimum order size to show that only modifying this portion of the profit function, as in the Insurance Experiment, is sufficient to identify risk aversion.

matter for reasons other than the continuation value of learning, then a change in Γ may still affect behavior when returns are guaranteed.

A final and more direct test of learning relates to the persistence of effects of a one-time return policy. If firms learn about demand, then receiving returns once should cause them to experiment and update beliefs, affecting their future stocking behavior. A risk neutral firm would stock less in subsequent periods unless expectations positively updated as the value of information is lower. By contrast, a sufficiently risk averse firm would increase stocking as learning reduces risk.

Proposition 2 (Learning about demand) *If firms face uncertain demand for a new product that they can overcome through learning*
a.) New product stocking increases when the firm is offered returns.
b.) Reducing future supply chain uncertainty, Γ, increases stocking.
c.) If $\delta \approx 1$, non-stock costs of helmet sales are low, and capital constraints are not binding, then a reduction in supply chain uncertainty has no effect on $\mathbb{I}(I^*_{nt} > 0)$ if a firm has access to returns.
d.) A one-time return offer will have persistent effects on new product stocking.

Suppose that a one-time returns policy has a persistent positive effect on stocking. This is consistent with risk aversion flipping the relationship between demand uncertainty and new product adoption, deterring firms without access to a form of insurance from stocking the new product when they are uncertain about demand. Under this explanation, returns should cause firms with uncertain priors to stock, who then overcome uncertainty about demand and stay in the market. An alternative explanation is that returns cause firms with incorrect and pessimistic expectations to try stocking helmets, who then correct their beliefs. The model predicts a different relationship between beliefs and product adoption under these mechanisms.

Proposition 3 (Learning mechanism) *Suppose that a one-time returns policy has a persistent positive effect on new product adoption.*
a.) If returns correct the pessimistic priors of risk neutral firms, each of the following should hold:
- *i.) Prior uncertainty about demand and uptake are positively correlated absent returns.*
- *ii.) Returns cause firms with pessimistic priors to stock.*
- *iii.) Experience selling the new product causes firms to positively update beliefs.*

b.) If returns cause risk averse firms to stock and overcome demand uncertainty:
- *i.) Prior uncertainty about demand and uptake are negatively correlated absent returns.*
- *ii.) Returns cause firms with uncertain priors to stock.*
- *iii.) Experience reduces uncertainty about demand.*

If learning is consequential to firms' investment decisions, then a natural question is whether they can learn from each other. Information externalities may contribute to risk averse firms remaining in the market if they can free-ride off of competitors to compensate for their own lack of risk taking. If information spills over, then receipt of a signal from neighbors should increase the propensity of a firm to stock if they are risk averse since uncertainty is reduced, and the effects of returns or a relaxation of supply chain frictions should be smaller. Formally,

Proposition 4 (Information externalities) *If firms can learn about demand from competitors*
a.) The effects of a returns offer or change in Γ *are decreasing in* φ*.*
b.) Exposure to a seller increases I^*_{nt} *if agents are risk averse or have pessimistic priors.*
c.) Receipt of a signal x_{ot} *below an agent's expectation increases* I^*_{nt} *only if agents are risk averse.*

1.3 Setting and design of the experiments

I test the model's predictions using two field experiments in Kenya. First, an Insurance Experiment induces a mean-preserving contraction to test if firms are risk averse. Second, a Learning Experiment traces out the longer-run trajectory of helmet adoption by temporarily alleviating risk using a realistic policy, then examining the persistence of effects.

Both experiments offer a motorcycle helmet introduced 2.5 years before the study began. Previously, helmets affordable to a typical Kenyan household offered minimal safety benefits. However, a motorcycle importer invested in a modern helmet factory in Kenya in 2020. Lower transportation and labor costs brought more effective helmets into the budget set of consumers.[4]

This market provides an ideal setting to study new product adoption because retail adoption of helmets was slow despite a large base of motorcycle users. A census showed that more than two years after helmets were introduced, fewer than 3% of surveyed firms sold them. Yet motorcycle use is widespread: in 2022 an estimated 2.4 million motorcycles were operated each day, accounting for 22 million trips. Helmet use is limited, and the annual number of motorcycle deaths doubled between 2018 and 2021.[5]

The insurance experiment

Setting and timeline: The Insurance Experiment (November 2024 to April 2025) was conducted in 140 markets in West Kenya, spanning 9 counties. In each county, 15-16 large markets (identified by field officers) were selected. The experiment included a baseline survey followed by a follow-up 2 months later. A survey with a separate *spillover* sample was conducted after the follow-up.

[4] East Africa lacked testing labs at the time of the study, but the helmets used are considered high quality and were approved by the Kenyan Bureau of Standards. For details, see Strzyzynska, Weronika (2023). "Africa sees sharp rise in road traffic deaths as motorbike taxis boom." *The Guardian*.

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

The studied helmets are produced about 300 kilometers from this sample and were not widely available in West Kenya at the time of the experiment. Low-quality imports and alternative helmet types, such as construction helmets sometimes used by motorcycle riders, were more common. Approximately half of enterprises reported knowing of a nearby shop selling such products.

Recruitment and sample: The baseline survey included 70 randomly selected markets, stratified by county. The remaining 70 markets were pure controls visited only for the *spillover* survey.

In each baseline market, field officers identified 10 shops suitable for helmet sales, prioritizing motorcycle spare parts and repair shops. If fewer than 10 existed, other shops such as hardware stores were added. Eligibility required that shops did not currently sell helmets and had a fixed physical location. From the 10 eligible shops, 5 were randomly selected, yielding 350 firms.[6]

The focus on urban and semi-urban markets, plus the restriction that enterprises have a permanent location, resulted in a sample of relatively large and established firms. Median profits the month before the survey were about $365 (summary statistics are reported in Appendix Table A.1).[7] The average shopkeeper was 34 years old with 13 years of education and about 30% were female.

More than half of the shops sold motorcycle spare parts. General shops and hardware stores were the next most common categories. The average business age exceeded 4 years, and 1/4 of firms had an employee. Half of shop owners reported having no employees or other sources of family income. These enterprises may be sensitive to profit risks since they are the residual claimant to all firm revenues, which fund all household consumption.

The 5 shops not surveyed at baseline were targeted for the *spillover* survey. The aim was to offer these shops plus identically sampled firms in the control markets a stock of helmets to test for information externalities (Proposition 4). However, only the name of shops was recorded at baseline to prevent priming, making it challenging to confidently re-identify shops.

As a result, field officers instead targeted the 5 best shops to sell helmets in pure control markets. In baseline markets, they recruited the 5 best shops not already surveyed (including skipped firms).[8] Typically, all motorcycle parts shops in each market were recruited, and assignment to the *insurance* or *spillover* sample was determined randomly in baseline markets. Consequently, the sample of motorcycle-related shops should be comparable across

[6]One market was inadvertently visited in piloting and at baseline, resulting in some firms receiving a control and treatment offer. Analysis focuses on the 345 remaining firms, but results are similar including the full sample.

[7]Egger, Haushofer, Edward Miguel, Niehaus, et al. (2022a) report *annual* revenues of $240 among retail firms in rural Siaya County, Kenya.

[8]Some shops stocking helmets prior to the study, which were ineligible, reported adopting helmets after the study began to obtain subsidized stock. Back checks were conducted after the delivery of subsidized stock, so misreporting incentives were removed. Firms that disclosed eligibility violations in back checks are excluded from analysis.

baseline and pure control markets within the *spillover* survey. Because fewer motorcycle shops were available to survey in baseline markets (since some had already been visited), firms in these markets are generally less suitable to sell helmets. Appendix Table A.2 validates this, showing that *spillover* shops in baseline markets are less likely to sell motorcycle products, but shops are balanced across other dimensions.

Design and randomization: The Insurance Experiment was designed to test if risk aversion affects firms' stocking decisions (Proposition 1). The design randomly assigned firms to an *insurance* or control status, stratified by market. *Insurance* firms were offered helmet stock with a contract designed to induce a mean-preserving contraction of profits. Control firms were offered the same stock of helmets without this contract, allowing one to test if firms are risk averse by examining if *insurance* firms stock at a higher rate.

Both arms were offered a stock of 3 helmets at the end of the baseline survey and informed that they could purchase 3 or more helmets at a follow-up, unconditional of prior stocking. The order size (80% of median weekly profits) was set to the manufacturer's minimum. Shops paid full price but received free delivery. They were allowed two weeks to place an order, giving time to search for customers and raise capital.

Insurance shops were offered firm-specific contracts calibrated based on subjective beliefs about demand. To elicit beliefs, *insurance* and control firms were asked how many of 10 helmet-stocking shops would sell out by follow-up, and whether their own shop would earn more or less revenue than the average firm. Respondents earned small rewards for accurate predictions. Field officers then reported the implied probability of selling out, and each shopkeeper stated their own belief about this probability, so_i.

Control firms were promised an unconditional payment (in Kenyan shillings) at follow-up of

$$P_i = \begin{cases} 1000 \cdot (1 - so_i) + 25, & so_i < 0.8 \\ 225, & so_i \geq 0.8 \end{cases}$$

Insurance firms could choose P_i or a contract paying 1,000 if they stocked helmets and failed to sell out by the follow-up, but 0 otherwise. This exceeds P_i when helmet demand is low, mitigating losses. But the expected payout is less than P_i. Appendix A.1 shows that setting $P_i' = 1000 \cdot (1 - so_i)$ would make insurance a mean-preserving contraction. Since $P_i > P_i'$, *insurance* is strictly dominated under risk neutrality. Thus, higher stocking rates under insurance imply risk aversion.

The design also addressed confounding explanations. First, *insurance* shops were given the choice between a future transfer or insurance, ruling out any treatment effect due to capital constraints (Casaburi and Willis, 2018). Second, both *insurance* and control firms were informed of the contract offer and its random assignment, so that receipt did not provide a signal about demand. Third, firms accepting the payout could not restock, eliminating incentives to intentionally forgo sales (and further lowering the expected insurance payout). Finally, a mystery shopper visited 1 in 5 firms, buying a helmet if available (or another item otherwise). Firms refusing to sell helmets but reporting unsold inventory were denied payouts and barred from restocking.

The primary outcome is an indicator equal to 1 if a firm ordered helmets within two weeks. A secondary outcome captures “immediate adoption” (orders within 24 hours). The first measures adoption overall; the second isolates behavior before firms could mitigate risk by finding customers. Surveys also collected firm demographics, demand beliefs, realized sales, and risk preferences via a lottery-choice game conducted at the follow-up (Strobl, 2022). This exercise was omitted from the baseline to ensure that payouts did not confound effects of *insurance*.

Markets had 2 or 3 *insurance* firms with equal probability. Within markets, simple random assignment allocated *insurance* to firms. Appendix Table A.1 shows that balance across arms.

If no firm in a market ordered helmets, shops were unexpectedly approached again (in a random order) and offered a large discount, ensuring that each baseline market had a seller to test for information externalities using the *spillover* sample.

The learning experiment

Setting and timeline: The Learning Experiment (January-July 2024) was conducted in Nairobi, Kenya. A census identified eligible firms and collected data for randomization. Data collection included a baseline survey (February-March), a midline (May) and an endline (June-July). Chronologically, this exercise was conducted before the Insurance Experiment since the design relied on an environment where shops had prevailing access to stock.

A key feature of this setting is that the manufacturer is local, so helmets were available to firms for over two years. This allows one to test if varying exposure to risk changes stocking in a setting where product diffusion had been slow despite the producer’s efforts to promote their adoption.

Recruitment and sample: The Learning Experiment included 929 retail firms. Shops with prior experience selling motorcycle helmets were excluded since the study focuses on new product adoption. Recruitment focused on areas where few existing helmet shops were present. Consequently, recruitment was low in the Central Business District, where firms are large. Shops without a permanent building were excluded. Finally, shops that were certain they would never stock helmets were omitted to approximate the population of firms a producer may target when seeking retailers.

Recruitment occurred via a census: surveyors visited shops, showed them a sample helmet, explained eligibility, and recorded information for randomization. Of 1,152 eligible firms listed, 950-1,000 were targeted. Shops were placed in either a primary or replacement pool. The primary pool deterministically included firms located far from existing sellers, with others randomly assigned. Surveyors first recruited from the primary pool, drawing replacements if firms declined. The final sample included 929 firms: 22 interested shops failed to complete the baseline, and five were dropped after being deemed ineligible. Treatments were introduced after baseline, ensuring participation was not influenced by assignment.

Appendix table A.1 presents summary statistics. Firms in the Learning sample were larger than those in West Kenya, with average monthly profits over \$600, and a smaller share (37%) sold motorcycle parts. Respondent age, gender, and education levels were similar across samples, as were firms' likelihood of having an employee or adopting a new product in the prior year.

Returns treatment: The first arm was designed to test if temporarily reducing inventory risk led shops to permanently stock helmets (Proposition 2). This was carried out over two phases.

In phase one, all shops were offered a stock of helmets at prevailing market prices. *Returns* shops were given the option to return stock from their first order for a refund, whereas control shops could not. In phase two, both groups could restock without returns, making conditions identical. The key outcome is whether reduced risk in phase one raised stocking in phase two by resolving demand uncertainty. As in the Insurance Experiment, all firms were informed about the existence of treatments and their random assignment to avoid signaling.

Returns were used to reduce risk for two reasons. First, return policies are widespread in high-income settings (Padmanabhan and Png, 1995). This intervention therefore helps to address whether return markets are thin (less than 15% of *spillover* firms reported ever being offered returns).[9] Second, *returns* do not involve payouts that could affect phase two stocking.

During phase one, shops could order at least 5 helmets, half the manufacturer's minimum. This reduced liquidity needs and maintained goodwill among controls. Returns were permitted only once, at the midline, to target demand uncertainty rather than ongoing inventory risks. Phase two ran from midline to endline, with orders accepted anytime. Shortly after midline, the manufacturer unexpectedly cut its minimum order size to 3, and the study followed suit.

If a shop indicated intent to order, surveyors revealed an installment option: firms could pay for their first order in three weekly payments, regardless of treatment. This helped shops with limited liquidity while preserving time for learning about demand. To prevent contamination, surveyors disclosed this option only after a shop expressed intent to buy.[10] The default rate was under 3%.[11]

Supplier commitment treatment: The second treatment was designed to increase the continuation value of learning without affecting phase one helmet profitability or risks (Proposition 2).

I targeted uncertainty about future supply chain reliability to achieve this. Surveyors told *supplier commitment* shops that the study would help them restock directly with the manufacturer at endline, ensuring continued access if helmets proved profitable. Controls

[9]Anecdotally, the helmet producer cited the fact that contracting costs, mistrust, and shipping make returns costly when selling to many small retailers as barriers to offering returns.

[10]In three cases, the option was revealed early; those firms were dropped.

[11]91% of shops paid on time and 97% paid before the midline survey. Most shops that paid late received incorrect information about the repayment schedule and paid within 1-2 weeks of the intended time.

had to establish this link themselves. This design reflected pilot evidence that most shops did not know how to reach the supplier and feared that small retailers would be de-prioritized if the supplier's focus shifted. The focus of this treatment is on phase one uptake.

A secondary information treatment: The endline survey included a randomized *information* treatment designed to test if firms learn from peers (Proposition 4). *Information* shops received data from 5 randomly selected phase one helmet adopters about how many helmets they sold, the prices that they charged, and whether they restocked. Using 5 shops generated randomly varying signals: average sales were 3.8 helmets with a standard deviation of 1.2. This relied on trust built over prior surveys, making it a stylized test of information value and not a market intervention.

Randomization: Stratified random assignment was used to allocate the *returns* and *supplier commitment* treatments. I followed a 2x2 randomization design so that 1/4 of shops received *both* treatments to test predictions about the interaction between the offers (Proposition 2). Randomization was stratified on neighborhood (for power), an indicator for whether the enterprise reported having any employees during the census (for power), and distance to the nearest existing helmet seller (for heterogeneity analysis). Appendix Table A.1 verifies balance across arms.

The *information* treatment was delivered unexpectedly to half of shops that had not stocked helmets by the endline. Randomization was stratified on knowledge of a nearby helmet seller.

Primary variables collected

The baseline survey captured information about the demographics of the shopkeeper and the enterprise's costs, revenues, and profits. Field officers also elicited beliefs about helmet demand using a frequentist approach (Benjamin, Moore, and Rabin, 2017). Respondents first stated the price at which they would sell helmets, then were asked to imagine stocking 10 units. They received 20 beans and a sheet labeled 0–10, and were instructed to place beans across boxes to represent the probability of selling each quantity over a month.[12] From this distribution, both the expected value and the variance of beliefs can be calculated.

Reported expectations strongly aligned with a separate direct question about anticipated monthly sales ($t > 18, R^2 = 0.49$). Most respondents also provided plausible full distributions. A minority placed all beans on their expectation (or nearly all), yielding degenerate distributions that likely reflect measurement error. Results are stronger when excluding these cases, but they are retained in main specifications to be conservative. Measures of variance also face a censoring issue for very high expectations since boxes do not exceed 10. I elicited responses up to twice the Learning Experiment order size of 5 so that observations where censoring binds are unlikely to be affected by the *returns* arm.

[12] Field officers often entered totals that did not sum to 20 on the first day of data collection. Surveyors were retrained and the variable is set to missing for observations from day one.

The midline and endline surveys recorded helmet stocking from both the study and other suppliers. For adopters, they captured helmet sales, costs, revenues, and profits, as well as stocking intentions. All firms provided updated information on overall revenues, costs, and profits, plus new data on product offerings. Beliefs about helmet demand were re-elicited at each round.

Analysis focuses on phase one and phase two stocking decisions and on beliefs about helmet profitability. Helmet profits are also examined, but given that only a minority of firms stocked, the experiment was not powered to detect profit effects. For this reason, repeat stocking serves as the primary revealed-preference measure of profitability (this focus was pre-registered).

1.4 Empirical specifications

I examine the effects of the Insurance Experiment by estimating the regression

$$Stocked_i = \alpha + \beta Insurance_i + X_i'\gamma + \mu_m + \epsilon_i \tag{1.9}$$

$Stocked_i$ indicates that shop i stocked helmets, $Insurance_i$ denotes receipt of the *insurance* offer, X_i denotes controls, and μ_m market fixed effects. Proposition 1 predicts $\beta > 0$, stocking increases in response to the expected profit and variance lowering contract, only if firms are risk averse.

Turning to the Learning Experiment, I test if agents face demand uncertainty by estimating

$$Stocked_i = \alpha + \beta_1 R_i + \beta_2 S_i + \beta_3 R_i \times S_i + X_i'\gamma + \mu_k + \epsilon_i \tag{1.10}$$

$Stocked_i$ denotes phase one stocking, R_i indicates receipt of *returns*, S_i captures receipt of the *supplier commitment*, and μ_k is strata fixed effects. The interaction captures the model's prediction that the offers are substitutes. Proposition 2 yields four predictions: (a) $\beta_1 > 0$, since returns reduce loss risk; (b) $\beta_2 > 0$, since supplier commitment increases the continuation value of learning; and (c) $\beta_3 < 0$, since supplier commitment should have little additional effect when returns already eliminate risk. Part (d) predicts that *returns* will have persistent effects, reflecting direct evidence of learning. This is tested by replacing the outcome with phase two stocking and testing if $\beta_1 = 0$.

If returns have a positive effect on phase two stocking, Proposition 3 provides tests to differentiate between a case where *returns* overcome risk aversion versus incorrect beliefs. This is tested with three equations. First, I estimate

$$\begin{aligned} Stocked_i = \alpha &+ \beta_1 \mathbb{E}[Sales]_i + \beta_2 SD[Sales]_i + \beta_3 R_i \\ &+ \beta_4 R_i \times \mathbb{E}[Sales]_i + \beta_5 R_i \times SD[Sales]_i + X_i'\gamma + \epsilon_i \end{aligned} \tag{1.11}$$

$\mathbb{E}[Sales]_i$ is the agent's prior expectation about the number of helmet sales in the next month and $SD[Sales]_i$ is the standard deviation of their belief. Under the targeted mechanism (risk aversion), Proposition 3 predicts that $\beta_2 < 0$, meaning control firms with more uncertain beliefs are less likely to stock, versus $\beta_2 > 0$ if firms are risk neutral, reflecting the value of learning.[13]

Second, I examine the composition of adopters by estimating

$$Returns_i = \alpha + \beta_1 \mathbb{E}[Sales_i] + \beta_2 SD[Sales]_i + X_i'\gamma + \epsilon_i \tag{1.12}$$

among phase one adopters not receiving the *supplier commitment*. β_1 captures whether sales expectations differ among *returns* versus control adopters. Under risk neutrality, $\beta_1 < 0$, since returns should draw in firms with pessimistic priors. If firms are risk averse, β_1 need not be negative. Instead, $\beta_2 > 0$ would indicate that returns primarily induce firms with uncertain beliefs to adopt.

To examine how stocking affected beliefs, I estimate the two-stage least squares system

$$\begin{aligned}
\Delta asinh(\mathbb{E}[Sales]_i) &= \alpha_e + \beta_1 Stocked_i + \rho_e KS_i + X_i'\gamma_e + \mu_k + \epsilon_{ei} && \text{(1.13a)}\\
\Delta asinh(V[Sales]_i) &= \alpha_s + \beta_2 Stocked_i + \rho_s KS_i + X_i'\gamma_s + \mu_k + \epsilon_{si} && \text{(1.13b)}\\
Stocked_i &= \pi_0 + \pi_1 R_i + \pi_2 S_i + \pi_3 R_i \times S_i + \pi_4 KS_i + \pi_5 R_i \times KS_i \\
&\quad + \pi_6 S_i \times KS_i + \pi_7 R_i \times S_i \times KS_i + \nu_i && \text{(1.13c)}
\end{aligned}$$

$Stocked_i$ denotes phase one stocking. I control for knowing a seller at baseline, KS_i, and use interactions between KS_i and treatment as instruments for power. $\Delta asinh(\mathbb{E}[Sales]_i)$ is the change in the inverse hyperbolic sin of i's expected helmet sales from baseline to midline or endline. $\Delta asinh(V[Sales]_i)$ is the analogous change in i's variance of beliefs. If *returns* correct pessimistic priors, Proposition 3 predicts that $\beta_1 > 0$, indicating increased optimism about demand. It predicts that $\beta_2 < 0$ if *returns* mitigate risk aversion, meaning experience reduces uncertainty.

The final class of predictions address spillovers. I test (i) whether exposure to incumbent sellers substitutes for a firm's own experience, (ii) whether proximity to entrants raises adoption, and (iii) whether the information treatment corrects pessimistic priors or reduces uncertainty.

Let $TM_i = 1$(Market visited at *insurance* baseline). IT_i denotes receipt of the *information* treatment, HS_i equals one if the average sales presented in the *information* signal is greater than an agent's expectation, and μ are strata fixed effects. I test for information externalities by estimating

$$\begin{aligned}
Stocked_i &= \alpha_1 + \beta_1 R_i + \beta_2 S_i + \beta_3 R_i \times S_i + \beta_4 KS_i + \beta_5 KS_i \times R_i \\
&\quad + \beta_6 KS_i \times S_i + \beta_7 KS_i \times R_i \times S_i + X_i'\gamma_1 + \epsilon_{1i} && \text{(1.14a)}\\
Stocked_i &= \alpha_2 + \beta_8 TM_i + \mu_c + \epsilon_{3i} && \text{(1.14b)}\\
Stocked_i &= \alpha_3 + \beta_9 IT_i + \beta_{10} HS_i + X_i'\gamma_2 + \mu_n + \epsilon_{2i} && \text{(1.14c)}
\end{aligned}$$

[13] I estimate this model among the restricted sample of shops not receiving the supplier commitment since the focus is on untreated behavior and the effect of returns. Results are similar if the full sample is used.

Proposition 4 part (a) predicts that learning from incumbents lowers the value of returns and supplier commitment, implying $\beta_5 < 0$ and $\beta_6 < 0$. Part (b) states that externalities exist if exposure to a seller increases adoption, implying $\beta_8 > 0$. Finally, Equation 1.14c tests whether spillovers operate through expectations or uncertainty. If they correct pessimistic priors, a signal only raises stocking when it exceeds expectations ($\beta_9 < 0$, $\beta_{10} > 0$). If they mitigate uncertainty for risk averse firms, then even a slightly negative signal can increase stocking, implying $\beta_9 > 0$.

1.5 Results

Demand priors and baseline stocking

Before turning to experimental results, I present descriptive evidence from firm beliefs suggesting that risk aversion may constrain helmet stocking (Appendix Table A.3). I focus on the Learning sample, where helmets were offered at market prices. When asked about a hypothetical purchase of 10 helmets, the median firm reported a 30% chance that the investment would reduce overall profits, compared to a 50% chance that it would increase them. Frequentist measurements similarly indicate that 50% of control enterprises expected to earn enough revenue to pay for the cost of stocking 5 helmets by the midline survey. Appendix Table A.7, discussed later in this section, validates that beliefs predict realized outcomes, suggesting that this evidence of optimism among many firms is not likely a product of measurement error.

This indicates that about half of firms believed that helmets are profitable in expectation. However, beliefs are uncertain, and most firms reported a risk of helmets resulting in losses. Consistent with risk aversion, under 7% of control shops accepted stock when it was offered to them. I therefore turn to experimental variation to evaluate whether lowering risk affects stocking.

Firm risk aversion

The results of the Insurance Experiment (Equation 1.9) show that access to the insurance contract–which lowered expected profits while reducing stocking risk–substantially increased helmet adoption, implying a rejection of firm risk neutrality.

Insurance increased the rate of shops that acquired stock within 24 hours by 7 percentage points on a base of 5% (Table 1.1 Panel A, $p < .01$). Including shops that stocked later, the effect is about 10 percentage points (50%, $p = .02$). This shows that lowering profit variance (while reducing expected returns) substantially increases the propensity of firms to stock a new product, implying that risk aversion affects firms' investment decisions. Effects of *insurance* are more than twice as large among firms operated by more risk averse shopkeepers, consistent with a lack of separability between owners' consumption preferences and production decisions ($p = .009$, columns 5-6).

The rate and composition of *insurance* shops that opted into the contract suggest high comprehension of the offer. Half of *insurance* shops that immediately stocked helmets selected into the contract, while 40% of firms that ever stocked did (Table 1.1 Panel B). These rates align almost exactly with treatment effects on stocking. Shops with more uncertain beliefs were far more likely to choose the contract, consistent with risk aversion. A one helmet increase in a shop's standard deviation of sales was associated with a 24 percentage point increase in contract selection among immediate adopters (column 1), and a 15 percentage point increase among all adopters (bootstrapped $p = .049$, column 2).[14] The fact that shops with low uncertainty about demand typically declined the contract is consistent with shops understanding that expected payouts of the contract were lower than the guaranteed offer. Anecdotal reports also support this view: in nearly every survey, shops complained that *insurance* should be more generous, frequently objecting to the condition prohibiting restocking and collecting the payout.

A potential concern is that *insurance* could raise expected profits if beliefs were mismeasured. If this were driving the results, then contract acceptance would be more likely when the gap between the survey-estimated expected value of the *insurance* contract and guaranteed payment were smaller since less error would be required to flip the optimal choice. There is no such relationship in the data (Table 1.1, Panel B). Moreover, average ex-post payouts would have been higher among *insurance* adopters had they declined the contract. Thus, under both elicited beliefs and rational expectations, insurance is strictly dominated for risk neutral firms. Finally, beliefs about their probability of selling out strongly predict the outcome, with an elasticity of 0.2 among all baseline adopters ($p = .05$) and an elasticity of 0.4 among those that accepted the contract.

Treatment effects remain large and significant among larger and older firms, suggesting that results are not driven by new or struggling enterprises. *Insurance* increased immediate helmet adoption by 20p.p. ($p < .01$) among firms with employees and the rate of shops that ever stocked helmets by 25p.p. ($102\%, p = .02$, Table 1.2). Point estimates are also large and significant among firms with higher profits and older enterprises.[15] These results suggests that risk aversion constrains entrepreneurship even when enterprises have liquidity available to make investments. From a policy perspective, this may also reduce the returns to capital in firms, since enterprises may direct additional resources towards low risk and low return investments absent insurance.

The stronger response to *insurance* among larger firms and firms with more risk averse owners is consistent with effects operating through owners' preferences rather than capital constraints generating concave returns to investment. If stocking helmets crowded out other inventory or operating expenses, convex borrowing costs could cause risk neutral firms to favor safer investments to avoid negative shocks forcing costly borrowing. However, one would expect larger *insurance* effects among smaller firms if this explained results since they

[14] On average, the bad state insurance payout was about \$11 larger than the guaranteed payment among firms that accepted insurance, corresponding to about 25% of the price of the helmet stock.

[15] Control stocking rates do not differ substantially by firm size. This likely reflects shops that did not perceive risk from stocking and those that found customers before stocking.

have less working capital, the opposite of the observed results. Directly supporting the view that such concerns are slack among compliers, under 4% of shops that adopted helmets reported that they had to stock less of other goods to afford the stock. Finally, the effects of *insurance* remain significant, and are suggestively larger, among firms that reported above median baseline access to liquidity.[16]

The finding that effects are stronger among larger firms suggests that, to the extent that such firms do not face binding credit constraints, firm owners' preferences may exhibit prudence. However, even firms that were not liquidity constrained at the time of investment likely face a risk of binding future constraints, which is sufficient to generate distortions in stocking under quadratic utility. The inter-quartile range of monthly profits at follow-up relative to baseline is $0.625 - 1.93$, indicating substantial income volatility that makes future binding constraints plausible even among currently liquid firms. Given challenges with identifying a sample of firms facing no risk of future credit constraints, combined with the limited sample size of such firms, I do not attempt to distinguish whether prudence or buffer stock savings concerns drive the results.

Appendix A.2 estimates a simple model of the decision of firms with CRRA utility to stock helmets, identified using random assignment to *insurance*. For consistency with the lottery choice game, this approach considers only short-run payoffs, not changes to lifetime income. This assumption is plausible to the extent that agents have limited ability to smooth shocks across periods, but would underestimate the level of risk aversion if agents can self-insure. The model suggests a mean CRRA coefficient of 0.62, a coefficient of 0.53 among those with below-median lottery choice measured risk aversion, and a coefficient of 2.21 among firms owned by those with above-median lottery choice measured risk aversion. These estimates suggest that modest levels of risk aversion can consequentially affect business decisions. However, these results are only suggestive as the parameters are imprecisely estimated and depend on strong assumptions.

Learning about demand

The effects of *returns* and the *supplier commitment* on stocking in phases one and two of the Learning Experiment (Equation 1.10) match the predictions of Proposition 3, consistent with demand uncertainty and learning consequentially affecting helmet uptake.

Returns increased the rate of firms that stock helmets during phase one, either from the study or any other supplier, by 9 percentage points relative to a 6.8% control rate ($p < .01$, Table 1.3). The effect remains stable, also 9 percentage points ($p = .02$), when the dependent variable is an indicator for ever stocking helmets by the endline survey. This suggests that *returns* increased firms' willingness to experiment with selling helmets, rather than simply accelerating adoption among firms that would have stocked regardless.

[16]The coefficient on *insurance* $\times$ above median liquidity is 0.20 ($p = .062$) in a regression of uptake on *insurance*, high liquidity, their interaction, and controls for family income, beliefs about demand, and employee hours.

The *supplier commitment* increased phase one helmet stocking by about 6 percentage points ($p = .03$, Table 1.3), but only among firms without *returns*. Consistent with the model, the treatment had no effect among shops offered *returns*, in which case firms that perceive any chance of profitability should already adopt. This pattern supports the interpretation that firms internalize a "risk-reward" trade-off, increasing investment when the future value of learning is high, at the cost of short-run utility. The absence of effects from the *supplier commitment* conditional on *returns* suggests that alternative explanations – namely learning by doing or fixed costs – are unlikely to explain the results, since these forces would predict an effect regardless of *returns*.[17]

The null effect of the *supplier commitment* conditional on *returns* suggests that the firms moved by the treatments were unlikely facing binding capital constraints and that non-stock costs of selling helmets are low. Consistent with this, under 15% of shops in the Learning Experiment reported any fixed costs of selling helmets, accounting for under \$2.50 on average (Appendix Table A.8). The Insurance Experiment asked directly if firms incurred any costs to sell helmets other than buying stock. Under 3% of shops reported such costs, averaging \$4. Furthermore, shops did not typically report greater work hours or effort to sell helmets, consistent with slack capacity (M. W. Walker et al., 2024). The view that compliers did not face binding liquidity constraints when purchasing matches the result from the Insurance Experiment that effects were larger among bigger firms.

More strikingly, *returns* had large and significant persistent effects on stocking in phase two, when *returns* and control shops faced identical stocking conditions. *Returns* firms were 7 percentage points (70%, $p = .03$) more likely to stock in this phase (Table 1.3). In markets without a pre-existing helmet seller (Table A.4), this rises to 8 percentage points (105%, $p = .02$). *Returns* also had large effects on market entry, defined by restocking and reporting intent to continue selling helmets (Table 1.3). *Returns* effects doubled the number of permanent market entrants based on this definition.[18] These results suggest that agents learned about demand since the groups differ only in that *returns* shops had more experience selling helmets in phase one, matching Proposition 3.

The persistent effect of *returns* on stocking provides revealed preference evidence that many shops found helmets profitable, which matches descriptive data. On average, phase one adopters reported profits of \$70 from helmet sales by endline, with fewer than 7% reporting losses. 65% restocked, and over 80% reported intent to stay in the market (Figure 1.1). Shops sold about 10 helmets on average by the end of the study at a typical price 1.5 times the wholesale cost. By endline, shops that stocked helmets reported that they accounted for about 10% of total profits, and they expected the share to rise to 20%. Consistent with these patterns, very few *returns* shops (7% of those that stocked) returned any helmets. These were concentrated among firms that reported losses: 80% of returning firms reported losses, versus 3% of shops not making returns.

[17]The binary outcome could generate this result due to ceiling effects, but Appendix Figure A.1 shows that many shops interested in stocking did not make purchases.

[18]Results are robust to other entry definitions, including stated intent to sell helmets or stocking three or more times.

Does risk aversion prevent experimentation with uncertain new products?

These results suggest that *returns* helped unlock a growth opportunity for many firms and expanded consumer access to helmets. The findings are consistent with risk aversion inhibiting experimentation with products whose demand is uncertain, but could also be driven by *returns* increasing expected profits, correcting pessimistic expectations.

The results of Equations 1.11-1.13 indicate that *returns* caused risk averse firms to try stocking helmets who overcame demand uncertainty through learning. Each of the predictions of Proposition 3 is satisfied under this mechanism, while none of the predictions under the alternative of *returns* changing expectations hold. The results suggest that risk aversion can substantially inhibit new product adoption by deterring firms from experimenting with risky products, limiting firm growth.

First, the results of Equation 1.11 show that, absent *returns*, firms more uncertain about the profitability of helmets are significantly less likely to stock (Table 1.4). This is consistent with risk aversion undermining experimentation: firms uncertain about demand avoid stocking. Otherwise, one would expect the opposite, since the option value of learning rises with uncertainty. I estimate regressions of phase one stocking on $SD[Sales]_i$ plus this variable interacted with assignment to *returns*, controlling for expectations. As discussed in Section 1.3, $SD[Sales]_i$ is truncated when expected sales approach 10. Furthermore, if expectations are sufficiently high, $SD[Sales]_i$ is an imperfect proxy for uncertainty about the profitability of stocking 5 helmets.[19] I thus exclude cases where firms perceive no risk that 1 month helmet revenues will fail to exceed costs (25% of cases).

Estimates in column 1 of Table 1.4, which includes rich firm and respondent controls, indicate that when the standard deviation of beliefs about sales increased by one, firms' propensity to stock fell by 5pp (80%, $p = .03$). Appendix Table A.5 shows that this relationship is similar using beliefs about revenue rather than sales, excluding covariates, and selecting controls using double-post LASSO. Estimates are also qualitatively similar, albeit underpowered, when observations prone to censoring are kept. Dropping cases where respondents report degenerate belief distributions strengthens effects. One may construct an additional robustness check by leveraging the fact that the outcome is firms' binary decision to stock 5 helmets. If agents' beliefs are sufficiently high that they perceive no risk of a loss from acquiring 5 helmets, theory no longer predicts a negative relationship between accepting stock and the standard deviation of beliefs. Consistent with this, the standard deviation of sales positively predicts stocking among firms perceiving no loss risk (Appendix Table A.5, column 6), although the difference is not statistically significant ($p = .24$).

Second, the results of Equation 1.12 indicate that the composition of firms that stocked helmets with *returns* held more uncertain but similarly optimistic beliefs versus control firms at baseline (Appendix Table A.6). This is consistent with treatment compliers being

[19] Put differently, if a shop is uncertain about selling 5 helmets or 10, risk aversion may lead to under stocking, but should not prevent purchasing 5. So a censoring issue affects the binary outcome measure for high expectations.

constrained from stocking by demand uncertainty and risk aversion, not incorrect beliefs. The standard deviation of *returns* adopters beliefs about the number of sales they will make in the subsequent month is 0.2 units higher versus control shops (24%, bootstrapped $p = .069$). There is no significant difference in expectations. This compositional effect is also evident in Table 1.4, which shows a large and significant interaction between treatment assignment and belief uncertainty: *returns* disproportionately increased adoption among firms with uncertain priors.

Third, and most directly, there is no evidence that the expectations of phase one adopters positively updated, but there is clear evidence of reduced demand uncertainty. Table 1.5 reports the results of Equation 1.13, which examines the effect of stocking in phase one (instrumented for using treatment assignment) on the change in an agent's belief about expected sales or the variance of sales from baseline to later surveys. There is no significant evidence expectations positively updated, and the point estimate is negative at endline. In contrast, estimates indicate that stocking caused the variance of compliers' beliefs about demand to fall by over 70%[20] ($p = .032$). The reduced-form also supports these conclusions: *returns* is associated with a small and insignificant decrease in expected sales but a 0.145 unit reduction in posterior variance ($p = 0.077$).

The view that learning reduced uncertainty without substantially changing expectations is robust to alternative variable definitions and specifications.[21] Running the same IV specifications pooling survey waves on expected revenue, obtained by combining frequentist beliefs about sales with prices, indicates a small and insignificant effect on expected profits but a \$26 decrease in the standard deviation of revenue ($p = .05$). And regressing the change in expected sales or the standard deviation of sales on stocking in phase one (with no instrument) suggests adopters expect to make .03 fewer sales but the standard deviation of their belief is smaller by 0.19 helmets ($p = .04$).

Reductions in belief variance among phase one adopters translate into substantially improved predictive accuracy over time. This validates the belief measures and indicates that declines in reported uncertainty are not likely driven by experimenter demand effects or overconfidence. Appendix Table A.7 reports the elasticity of realized sales with respect to expectations at baseline and at midline, restricting the sample to shops that accepted stock in phase one. At baseline, firms held diffuse priors, so attenuation bias implies an elasticity below one. Consistent with this, the estimate is 0.28 ($p = .024$, $R^2 = .045$). At midline, firms report substantially less uncertainty, which should yield stronger predictive power. The elasticity rises sharply to 0.69 ($p < .001$, $R^2 = .239$), and one can reject the null hypothesis of no change predictive accuracy with 95% confidence ($p = .047$).

Direct tests of learning present a consistent picture: for *returns* compliers, experience selling helmets left expectations about demand essentially unchanged but substantially reduced the variance of beliefs. This suggests that effects of *returns* on phase two stocking reflect risk aversion preventing firms uncertain about demand from stocking, causing a learn-

[20] Pooled point estimate: $\exp(.13 - 1.45) - 1 \approx -.73$ where 0.13 is the intercept.

[21] These results are only reported in the text for brevity given their similarity to Table 1.5.

ing breakdown. Importantly, these results apply specifically to the subset of firms induced to try helmets by *returns*, who on average held more optimistic but uncertain beliefs. Some firms likely held inaccurate (pessimistic) beliefs, but the intervention was not strong enough to induce stocking among them.

Information externalities

The results of the Learning Experiment suggest that learning about demand is a consequential determinant of new product adoption. Because firms are risk averse, investing in new products involves high utility costs. Can firms learn from each other to avoid making risky investments themselves? The results of Equation 1.14 suggest that firms can substitute for learning from their own experiences by observing neighbors stock, validating Proposition 4.

Externalities matter for two reasons. First, they further support the argument that the persistence of *returns* reflects learning about demand since exposure to information from sources other than one's own experience generates similar effects. Second, information spillovers suggest a market imperfection that helps rationalize how risk averse firms survive competition. Absent externalities, one would expect more risk averse firms to have lower productivity and be driven out of business. But spillovers may allow risk averse firms to mimic neighbors product discoveries, rationalizing their survival (and potentially undercutting risk taking incentives of neighbors).

Treatment effect heterogeneity aligns with the model's predictions. Neither *returns* nor the *supplier commitment* affected stocking when there was a pre-existing helmet seller near the firm, and one can reject the equality of effects of *returns* with 95% confidence (Table A.4). The adoption rate of untreated firms near a pre-existing seller is also approximately equal to that of *returns* firms in markets without incumbent motorcycle shops (Table A.4).

Results of the *spillover* survey, presented in Table 1.7, experimentally validate the view that firms learn from each other. The coefficients on "BL market" reports the effect of being in a market where a shop was induced to stock helmets three months before the survey. Results are reported over motorcycle-related shops and all shops in the sample. As detailed in the design, the sample of motorcycle shops is likely balanced across markets, whereas across the full sample firms in baseline markets were ex-ante less likely to stock.[22]

Firms observed competitors adopt helmets, leading to an increase in their own propensity to stock. Across the sample of motorcycle shops, firms were about 13 percentage points more likely to report knowing a helmet seller in their market ($p = .03$) and to purchase stock ($p < .01$). Across all shops, the estimated effect on stocking falls to 6 percentage points but remains statistically significant ($p < .01$). Anecdotal reports from shops help shed light on how information spillovers occurred. Several firms reported that customers entered their shops with pictures of the study helmets and asked if they could provide them at a better

[22]Validating this argument, surveyed shops in baseline markets are 20 percentage points less likely to sell motorcycle parts, and motorcycle-related shops are more than four times as likely to stock helmets in pure control markets.

price. Appendix Table A.9 shows similar results using non-random variation in the Learning Experiment.

I further validate that information matters directly by examining the effects of the *information* treatment (Table 1.6). Receipt of the anonymous sales data increased the rate of shops that stocked helmets from 0.6% to 2.8% ($p = .02$). As with *returns*, the evidence suggests that information primarily affected behavior by reducing uncertainty. When a "high signal" indicator – equal to 1 if the data the shop received exceeded their expectation – is added (Equation 1.14c), the effect of receiving any data, whether above or below expectations, remains large and significant.[23]

While spillovers were strong, evidence suggests that they were localized, explaining slow helmet diffusion in Nairobi. In the *spillover* sample, only 42% of shops in treated markets reported knowing of a local helmet seller. In the Learning Experiment, shops within 0.25km (about 3 blocks) of a study shop that adopted helmets were 14p.p. more likely to report knowing of a seller ($p < .01$), but the presence of a shop within 0.25-0.5 km had no effect (Appendix Table A.9). This suggests minimal spillovers across markets, preventing social learning from ensuring diffusion.

1.6 Discussion

The results indicate that inducing experimentation with helmets led to large and permanent increases in stocking. Could the persistent effect of *returns* on helmet stocking be driven by a mechanism other than learning resolving demand uncertainty?

One possibility is that *returns* increased phase one profits, relaxing capital constraints. But in practice, *returns* had no significant effect on overall firm profits because high profit volatility dominated reported gains. Stocking effects also survived the fall in order size from 5 to 3 helmets, which lowered liquidity requirements for all firms. Direct evidence of learning and information spillovers further support the view that persistence reflects firms overcoming uncertainty. Lastly, flooding impacted about 1/5 of shops at the end of phase one, resulting in a 20% ($96) profit decline among exposed firms (Appendix Figure A.2, $p = .024$). Despite this shock, phase two stocking among *returns* shops affected by flooding exceeded that of unaffected control firms ($p = .021$).

A second potential mechanism is learning by doing. However, helmet profits between midline and endline were not higher than those between baseline and midline among adopters, or higher among shops that stocked helmets earlier (Appendix Table A.8). The lack of effects of the *supplier commitment* given *returns* and null treatment effects in markets with a pre-existing seller also provide evidence against this explanation. And finally, the fact the information about demand from sources other than one's own experience (neighbors or the

[23] This leverages the fact that signals were constructed from 5 random shops. The mean signal was 3.8 sales (SD 1.18). In columns 2 and 4, "high signal" is constructed by comparing the signal to the agent's elicited expectation about demand, controlling for expectations. In column 3 and 5, I examine whether the signal exceeds the median.

study) increases stocking suggests that learning about helmet profitability, not learning by doing, is the primary mechanism.

Third, behavioral biases such as loss aversion or cognitive uncertainty may make *insurance* or *returns* valuable for reasons distinct from neoclassical risk aversion. The central argument of this paper is that small firms' production decisions depend on owners' risk preferences. This still holds, and in fact may be more important, if behavioral biases affect consumption preferences. However, the evidence is more consistent with standard models of risk aversion. The response to the *supplier commitment* intervention rules out preferences like maxmin utility since investment expands when the future value of learning increases, with no change in risk. Moreover, appendix A.2 suggests that modest levels of risk aversion can rationalize the observed response to *insurance*.

The large effect of *returns* on firm entry at endline suggests that the Learning Experiment was effective at expanding consumers' access to motorcycle helmets, an important safety product. This market expansion also suggests that risk aversion can reduce the market size that producers face when introducing new goods, harming incentives to innovate. Smoothing risk aversion, such as by facilitating returns, could be an effective tool for practitioners that aim to strengthen markets for new products. However, there are two concerns with the results that affect the policy interpretation. First, did the intervention expand helmet access or displace economic activity, crowding out firms that otherwise would have began stocking helmets? Second, would all of the *returns* firms that began selling helmets have adopted them anyways once they observed peers stock?

Information spillovers indicate that displacement is unlikely, suggesting that if anything the entry of *returns* shops made competitors more likely to enter the market. I also test for displacement in columns 1-2 of Appendix Table A.10. The dependent variable in column 1 equals 1 if the surveyed shop was stocking helmets at endline or reported that a shop near them was.[24] *Returns* increased the likelihood that the shop or an enterprise near them sold helmets by about 8 percentage points (33%, $p = .04$). The effect is larger if I examine whether a seller was ever reported near the sampled shop: *returns* had about a 13 percentage point effect (42%, $p = .03$). These results are consistent with the finding that effects are concentrated in markets with no baseline seller, suggesting that the intervention crowded in firms in areas where there were no peers to learn from.

The question of whether *returns* entrants would have eventually stocked absent the treatment is more challenging to answer since the study ended after six months. However, helmets were available to shops for over two years before the study, and the manufacturer made a strong effort to market to shops. This suggests that, at least in the near term, shops were unlikely to enter absent the intervention. The effects on helmet sales during the study period are also economically important even if the study did not affect long-run helmet access. Appendix Table A.10 suggests that *returns* induced shops to stock about 1.9 additional helmets on average during the study and sell about 1.4 more (144%, $p = .07$). In markets with

[24]This approach relies on shops to determine what firms they consider to be proximate. Field officers' assessment of the presence of shops and those of respondents were highly correlated at baseline.

no baseline helmet seller, these point estimates jump to 2.8 additional helmets stocked and 2.1 sold (528%, $p = .03$). These values correspond to *returns* inducing over 600 additional helmet sales in this time, accounting for around $15,000 in sales.

1.7 Conclusion

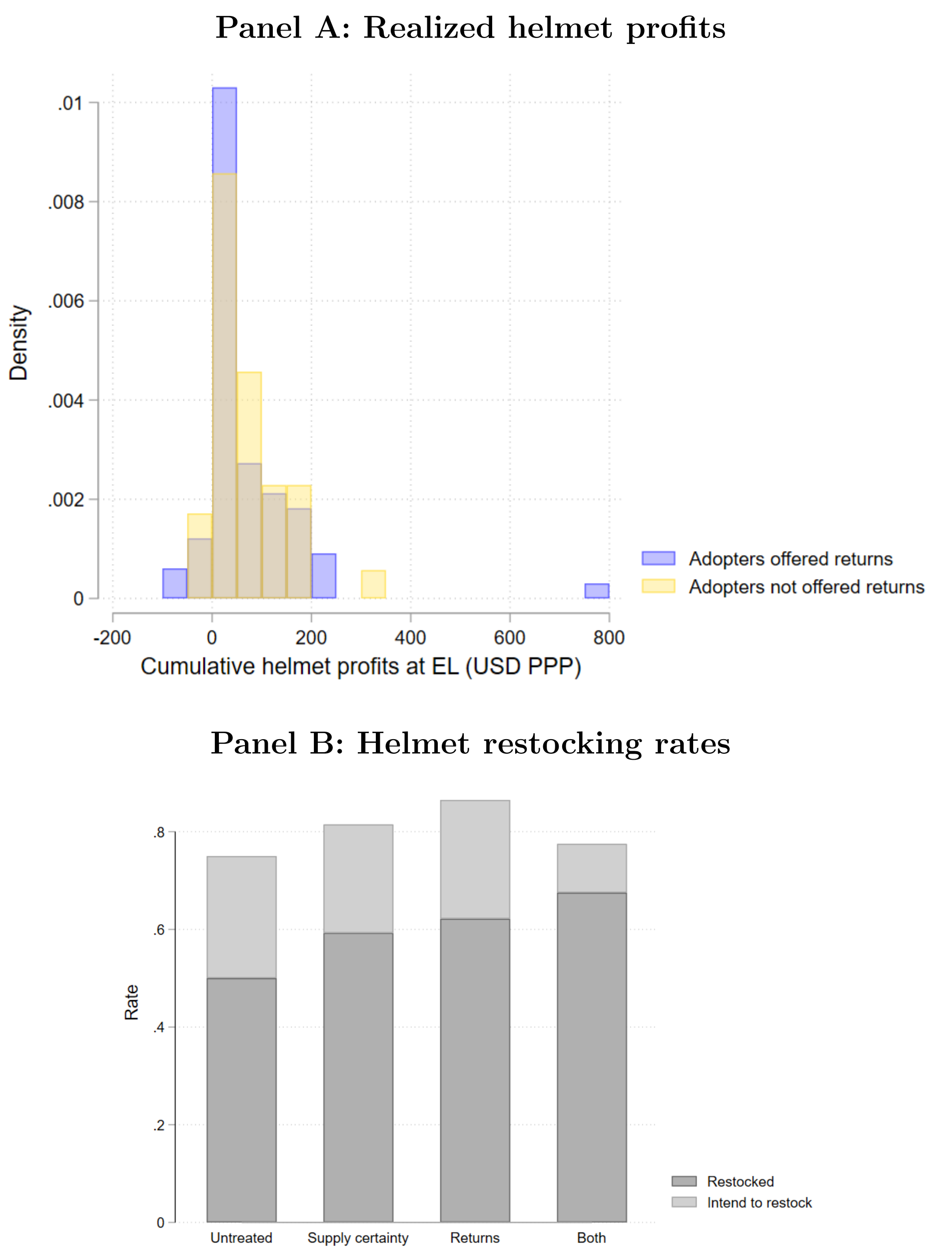
This paper provides evidence that risk aversion among small retail firms in Kenya acts as a binding constraint on the adoption of a profitable new product. I first find that offering firms an insurance contract that lowered both risk and expected profits led to a 50% increase in product adoption. This approach yields a robust test for risk aversion that can be readily adapted to other settings. I then investigate the dynamic implications of risk aversion by testing a model of firm learning. I find that temporarily mitigating inventory risk generated an increase in product stocking after the policy ended. This suggests that risk aversion prevents firms from engaging in experimentation needed to learn about new product demand. Finally, consistent with bandit models of learning, exogenously raising the future value of information increased upfront experimentation.

The substantial expansion in product stocking generated by these interventions suggests that risk aversion can constrain the retail market for new products in LMICs. Beyond slowing retailer growth, these distortions may harm incentives for upstream producers to introduce novel goods or invest in context-specific innovations. While mechanisms like return policies can mitigate retailer risk aversion, the logistical, transportation, and contracting costs of these offers may be prohibitive when selling to many small retailers. Consequently, the prevalence of small retailers in developing markets may disadvantage innovative producers. Future research investigating how the structure of LMIC markets contributes to incomplete financial contracts and studying policies to counteract risk aversion may therefore be valuable to promote product diffusion and innovation in LMICs.

More broadly, the finding that small firms are risk averse calls into question many economic models of small firms and policies designed to promote their growth. These results have broad relevance as over half of workers in LMICs are self-employed (International Labour Organization, 2019). Risk aversion may help rationalize slow technology adoption (Cirera, Comin, and Cruz, 2022), inefficient location choices (Pelnik, 2024), and foregone investments with positive net returns (Mel, McKenzie, and Woodruff, 2008). It may also lower the effectiveness of capital transfers if risk averse firms direct capital towards low-risk, low-return investments. Investigating the role of risk aversion in these choices, and the effectiveness of insurance policies, could be valuable to better align academic understanding of LMIC firms with their behavior and to identify effective policies to promote growth.

1.8 Tables and Figures

Figure 1.1: Learning experiment: Realized helmet profits and restocking rates



Panel A reports realized helmet profits at endline among firms that adopted helmets within a month of the baseline survey, breaking down the sample by those that received access to returns versus not. Profit estimates are net of lost profits on items shops were unable to stock to afford helmets. Panel B reports rates of restocking among the same set of shops. The darker bar indicates that the shop purchased at least 1 additional stock of helmets by endline and reported an intent to stay in the market, and the lighter bar denotes shops that had yet to restock but reported planning to.

Table 1.1: Insurance experiment: Helmet stocking and contract uptake

Panel A: Effects of insurance offer on helmet stocking

	(1)	(2)	(3)	(4)	(5)	(6)
	Stocked ($\leq 24H$)	Stocked	Stocked ($\leq 24H$)	Stocked	Stocked ($\leq 24H$)	Stocked
Offered insurance	0.075***	0.092**	0.073***	0.099**	0.002	-0.028
	(0.029)	(0.044)	(0.028)	(0.043)	(0.046)	(0.066)
High risk aversion					-0.066	-0.171***
					(0.045)	(0.065)
Offered insurance × High RA					0.168**	0.267***
					(0.075)	(0.102)
Observations	345	345	345	345	327	327
Control mean	0.047	0.203	0.047	0.203	0.047	0.203
Controls	Market FE	Market FE	Yes	Yes	Market FE	Market FE

Panel B: Insurance offer take-up and expected foregone returns among treated adopters

	Insurance uptake		Insurance expected payout − guaranteed		Insurance realized payout − guaranteed	
	(1)	(2)	(3)	(4)	(5)	(6)
	Stocked ($\leq 24H$)	Stocked	Dollars	Share	Dollars	Share
$\mathbb{E}$[Sales]	-0.017	-0.033				
	(0.113)	(0.036)				
SD[Sales]	0.239	0.153**				
	(0.221)	(0.078)				
Accepted insurance			-0.024	-0.003	0.932	-0.044
			(0.274)	(0.063)	(2.748)	(0.266)
Constant	0.224	0.290*	-0.822***	-0.117***	-6.046***	-0.472**
	(0.254)	(0.152)	(0.158)	(0.038)	(1.983)	(0.213)
Observations	22	51	51	51	47	47
Mean	0.500	0.392	-0.812	-0.113	-6.048	-0.493

$* \ p < 0.1$, $** \ p < .05$ $*** \ p < .01$. Robust standard errors in parenthesis in Panel A. Bootstrapped standard errors in parenthesis in Panel B (constructed with 500 draws).

This table reports results of the Insurance Experiment. In columns (1), (3) and (5) the outcome is 1 if a shop stocked helmets within 24 hours of the survey. Columns (2), (4) and (6) report effects on stocking within 2 weeks. Columns (3) and (4) include controls for industry, revenue, days open/week, knowledge of a nearby helmet seller, indicators for helmet storage space, selling multiple products, and stocking a new product in the past year. High risk aversion indicates that the agent exhibited above median risk aversion in a lottery choice game. Panel B reports (endogenous) uptake of the insurance offer among treated firms that stocked within 24 hours in columns (1) and within 2 weeks in (2). Columns (3) - (6) examine *insurance* shops that ever stocked helmets. Column (3) reports the expected value of the insurance offer less the guaranteed payment offered to firms, and column (4) reports column (3) normalized by the size of the guaranteed payment. Columns (5) and (6) are similar but use realized payouts rather than expectations. One market visited in both piloting and at baseline due to a sampling error is excluded from Panel A.

Table 1.2: Insurance experiment: Heterogeneity by firm size and age

	(1)	(2)	(3)	(4)	(5)	(6)
	Outcome: Stocked $\leq$ 24 hours			Outcome: Stocked		
	Employees	Firm profits	Firm age	Employees	Firm profits	Firm age
Treatment effect of insurance: smaller/younger firms						
Below median ×	0.025	0.020	0.036	0.046	0.039	0.043
insurance	(0.032)	(0.045)	(0.034)	(0.047)	(0.060)	(0.060)
Treatment effect of insurance: larger/older firms						
Above median ×	0.197***	0.118**	0.120**	0.249**	0.173***	0.170**
insurance	(0.071)	(0.047)	(0.054)	(0.100)	(0.066)	(0.066)
Pr(below = above)	0.039	0.160	0.219	0.070	0.130	0.169
Control mean ($\leq$ med)	0.044	0.072	0.033	0.193	0.193	0.228
Control mean (> med)	0.054	0.027	0.062	0.243	0.216	0.175
Observations	344	316	345	344	316	345

$* \ p < 0.1, \ \ ** \ p < .05 \ \ *** \ p < .01$. Robust standard errors in parenthesis.

This table reports heterogeneous treatment effects of *insurance* with respect to firm size and age. Panel A presents treatment effects of receiving an insurance offer on helmet stocking. In columns (1) - (3) the dependent variable equals 1 if the shop stocked helmets within 24 hours of the survey, while in columns (4) - (6) the dependent variable captures stocking within the 2 week window provided after the baseline. The below median group of firms by employee size contains no paid employees, while the above median group includes all firms with at least 1. Median firm profits in the last month were PPP USD 325. The median enterprise age is 3 years. Pr(below=above) reports the p-value of the test that the treatment effects are equal. Control means capture the untreated average of the dependent variable within the respective groups. All estimates include market fixed effects and controls for industry, baseline revenue, days open per week, knowledge of a nearby helmet seller, and indicators for having space to store helmets, selling multiple products, and stocking a new product in the past year.

Table 1.3: Learning experiment: Treatment effects on stocking and entry

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Stocked Phase 1		Ever Stocked		Stocked Phase 2		Entrant	
Returns	0.096***	0.089***	0.085**	0.078**	0.070**	0.065**	0.069**	0.066**
	(0.029)	(0.028)	(0.034)	(0.032)	(0.031)	(0.030)	(0.028)	(0.027)
Commitment	0.053*	0.060**	0.038	0.045	0.041	0.044	0.022	0.027
	(0.027)	(0.027)	(0.033)	(0.031)	(0.030)	(0.028)	(0.025)	(0.024)
Returns ×	-0.056	-0.042	-0.061	-0.043	-0.063	-0.047	-0.032	-0.021
Commitment	(0.043)	(0.042)	(0.049)	(0.046)	(0.044)	(0.042)	(0.039)	(0.038)
Observations	929	929	929	929	929	929	929	929
Control mean	0.068	0.068	0.131	0.131	0.102	0.102	0.068	0.068
Controls	Strata FE	Yes	Strata FE	Yes	Strata FE	Yes	Strata FE	Yes

$* \ p < 0.1, \quad ** \ p < .05, \quad ***p < .01$

This table plots treatment effects on helmet stocking and helmet market entry from the Learning Experiment. Column (1)-(2) report effects on phase 1 stocking, supplied by the study or any other source. Columns (3)-(4) examine effects on stocking at any point prior to the endline survey. Columns (5)-(6) examine effects on phase 2 stocking. Column (7)-(8) examine effects on entry into the helmet market, defined by two or more stock purchases and reported intent to permanently stock helmets. The dependent variable in columns (5)-(8) is coded to 0 if a shop withdrew from the study because they did not wish to sell helmets or if the outcome was confirmed without the survey. All estimates include strata fixed effects. Columns where controls indicates "Yes" also include industry fixed effects, controls for distance to the manufacturer, log revenue, days open per week, and indicators for stocking a new product in the year before the baseline survey, selling multiple products at baseline, and having space to store helmets without stocking less of another item.

Table 1.4: Learning experiment: Relationship between beliefs and adoption

	SD Sales			log(1 + Var sales)		
	(1) LPM	(2) Logit	(3) LPM	(4) LPM	(5) Logit	(6) LPM
Returns	-0.032	0.290	0.232	-0.042	0.947	0.243
	(0.077)	(1.339)	(0.185)	(0.117)	(2.207)	(0.184)
E[sales]	0.021	0.479		0.069	1.906	
	(0.015)	(0.311)		(0.046)	(1.208)	
$\sigma(sales)$	-0.052**	-0.933**	-0.053**	-0.061**	-1.127**	-0.067**
	(0.023)	(0.406)	(0.024)	(0.027)	(0.506)	(0.030)
Returns × E[sales]	0.013	-0.145		0.038	-0.743	
	(0.034)	(0.394)		(0.110)	(1.556)	
Returns × $\sigma(sales)$	0.058	0.984*	0.096**	0.079	1.274*	0.121**
	(0.055)	(0.586)	(0.048)	(0.064)	(0.721)	(0.059)
Observations	304	271	304	304	271	304
Control mean	0.068	0.068	0.068	0.068	0.068	0.068
Controls	Full	Full	Full	Full	Full	Full
$\mathbb{E}$[sales]× Returns FEs	No	No	Yes	No	No	Yes

$* \ p < 0.1, \quad ** \ p < .05, \quad ***p < .01$

Robust standard errors in parenthesis. The dependent variable in all regressions is an indicator equal to 1 if the firm stocked helmets in the month after the baseline survey. In columns (1) - (3), $\sigma(sales)$ is the standard deviation of the agent's beliefs about helmet sales, and in columns (4) - (6) this value denotes the log of 1 plus the variance of sales. The sample excludes firms offered *supplier commitment*. The sample further excludes shops with 0 loss probability from stocking 5 helmets to avoid top censoring. Estimates include controls for storage space, education, baseline profits, respondent characteristics, and firm characteristics.

Table 1.5: Learning experiment: Instrumental variable estimates of entry on posterior beliefs

	Instrumental Variables Estimates						Reduced form	
	(1) $\Delta\,\mathbb{E}[Sales]$	(2) $\Delta V[Sales]$	(3) $\Delta\,\mathbb{E}[Sales]$	(4) $\Delta V[Sales]$	(5) $\Delta\,\mathbb{E}[Sales]$	(6) $\Delta V[Sales]$	(7) $\Delta\,\mathbb{E}[Sales]$	(8) $\Delta V[Sales]$
Stocked (phase 1)	0.217	-1.173**	-0.026	-0.999*	0.095	-1.098**		
	(0.382)	(0.582)	(0.387)	(0.530)	(0.346)	(0.496)		
Returns							-0.001	-0.108*
							(0.047)	(0.063)
Supplier commitment							0.036	-0.040
							(0.043)	(0.058)
Returns × Supplier commitment							0.056	0.089
							(0.064)	(0.089)
Observations	820	820	832	832	1,652	1,652	1,652	1,652
Dep. var. mean	-0.111	0.059	-0.100	0.130	-0.069	0.062	-0.091	0.090
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weak IV robust p-val	0.561	0.050	0.859	0.073	0.794	0.032		
Survey	ML	ML	EL	EL	Pooled	Pooled	Pooled	Pooled

$* \; p < 0.1, \;\; ** \; p < .05, \;\; *** p < .01$

Robust standard errors in parenthesis in columns (1) - (4). Standard errors clustered by firm in columns (5) - (8). The dependent variables are the change in the log of 1 plus the moment of beliefs of firm i since baseline, for instance $\log(1+\mathbb{E}[\text{Sales at midline}]) - \log(1+\mathbb{E}[\text{Sales at baseline}])$. Columns (1) - (6) instrument for stocking in the month following baseline using treatment assignment and treatment interacted with an indicator for a seller at baseline, controlling for the presence of a seller. The baseline seller interaction absorbs heterogeneity in the first stage, yielding better mean-squared error (Abadie, Gu, and Shen, 2023). In columns (1) - (2), midline data is used, in columns (3) - (4) endline data is used, and in columns (5) - (8) both surveyed are considered. All estimates include controls for the saturation of baseline helmet sellers in the market, firms' reported willingness to take risks to experiment, the firm owner's performance on a digit span recall score, an indicator for whether the firm sold multiple products at baseline, and the firm's proximity to the helmet manufacturer. These controls were selected because they are likely to affect access to and retention of information affecting learning. Without covariates, the endline coefficient on expectations is $-.015 (p = .969)$ and the coefficient on variance is $-.87 (p = .096)$. Using CHS post-lasso-orthogonalized controls, selected from a list including those selected in the main specification as well as a wider list, yields a coefficient on endline expectations of $.004 (p = .990)$ and one on endline variance of $-.86 (p = .098)$.

Table 1.6: Information treatment: Effects on helmet uptake

		Helmet sales		Helmet revenue	
	(1)	(2)	(3)	(4)	(5)
1(Information treatment)	0.022**	0.030**	0.028**	0.029*	0.039**
	(0.009)	(0.015)	(0.014)	(0.017)	(0.017)
Signal > Expectation		-0.014		-0.011	
		(0.021)		(0.023)	
Signal > Median			-0.011		-0.030
			(0.017)		(0.018)
Expectation		-0.000		0.001	
		(0.000)		(0.003)	
Observations	727	722	727	722	727
Control mean	0.006	0.006	0.006	0.006	0.006

$* \, p < 0.1, \quad ** \, p < .05, \quad ***p < .01$

Robust standard errors in parenthesis. The dependent variable is an indicator for ordering helmets at the endline survey and the sample is restricted to shops that had not stocked helmets prior to that point. 1(Information treatment) indicates that a shop received helmet sale and price data from 5 randomly selected shops. Signal > expectation denotes that the signal average sales or revenue value exceeded the respondent's beliefs and is 0 otherwise or if the shop received no information, and Signal > median is one if the average sales or revenue value was greater than the median value across signals and 0 otherwise or if the shop received no signal. Columns (2) - (3) consider signals about helmet sales, and columns (4) - (5) examine signals about helmet revenue. All estimates include controls for knowing of a helmet seller at midline, log revenue in the month before the survey, and an indicator equal to 1 if the shop was affected by floods that occurred near midline.

Table 1.7: Spillover survey: Effect of helmet entrant on neighbor adoption

	(1)	(2)	(3)	(4)
	Know seller in market		Ordered helmets	
Panel A: Motorcycle shops				
Baseline (treated) market	0.115*	0.133**	0.098***	0.104***
	(0.069)	(0.061)	(0.035)	(0.036)
Observations	376	376	376	376
Control mean	0.361	0.361	0.091	0.091
Controls	No	Yes	No	Yes
Panel B: All shops				
Baseline (treated) market	0.122**	0.143***	0.048**	0.059***
	(0.052)	(0.049)	(0.022)	(0.022)
Observations	665	665	665	665
Control mean	0.300	0.300	0.070	0.070
Controls	No	Yes	No	Yes

$* \ p < 0.1$, $** \ p < .05$, $*** p < .01$

Standard errors in parenthesis clustered by market. Panel a restricts the sample to shops selling motorcycle-related products and panel b includes all firms. The variable "Baseline (treated) market" equals 1 if the market was randomly selected for surveys at baseline, and is 0 if the market was a pure control and was skipped. All results are from the spillover survey with non-baseline shops, collected 3-4 months after baseline. The dependent variable in columns (1)-(2) equals 1 if the shop reported knowing a helmet seller in their market (selling the studies variety, or any other variety). The dependent variable in columns (3)-(4) equals 1 if the shop purchased helmet stock. All estimates include county fixed effects. Estimates with controls include additional covariates for revenue, business age, employees, and the number of products the firm sells. All estimates exclude 35 observations from shops that started selling helmets before the study began. These shops were ineligible put initially misreported their sales start date to obtain subsidized stock. Back checks were conducted after delivery (when no misreporting incentive is present) with all shops selling helmets at the time of the survey to confirm eligibility. Shops that revealed they were ineligible during these back checks are excluded.

Chapter 2

Estimating Demand for Non-market Goods

2.1 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 their subjective beliefs without error. The primary identifying assumption is that measurement error in reported expectations (the gap between an agent's measured and true subjective belief) 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 patterns of belief misreporting such as rounding or overstating low probabilities. 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 healthcare, environmental regulation and transportation. Policymakers directly use VSL in benefit-cost analyses, and NGOs rely on the parameter to allocate development aid. 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 proxies for beliefs. This approach is biased if agents have beliefs about risk that differ from statistical estimates (i.e. their beliefs violate rational expectations). Moreover, estimates may be prone to omitted variable bias if mortality risk is correlated with other determinants of utility (Ashenfelter and Greenstone, 2004). Many studies also estimate VSL over highly selected populations where there is quasi-random variation in risk. Consistent with potential bias, estimates of VSL vary by orders of magnitude: in Africa, they range from \$0-\$700,000 (Berry, Fischer, and Guiteras, 2020; León and Edward Miguel, 2017a).[1]

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 unsure 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, Degroot, and Marschak (BDM, 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 over 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.

[1] 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.

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 2022 PPP USD $523.[2] 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 recovers informative bounds on VSL, rejecting values below $214 and above $1,043 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 Edward Miguel, 2017a), among a sample with incomes near $75,000, can be fully explained by typical macroeconomic estimates of utility curvature with wealth.[3] This aligns with evidence that health spending, as a share of consumption, rises sharply with income (Hall and Jones, 2007). I also estimate heterogeneity that matches theory: VSL is $500 higher among wealthier individuals in the sample, with an estimated income elasticity of at least 1.14, and higher among healthier respondents.

Is it plausible that the VSL in this sample is below annual income? This finding aligns with evidence of low demand for preventative health in LMICs. For instance, Cohen and Pascaline Dupas (2010) find that only 40% of pregnant women purchase malaria nets at a price of $0.6, implying a cost per life saved under $200 (Pryce, Richardson, and Lengeler, 2018). Similarly, Banerjee et al. (2010) show that $4 in-kind incentives substantially increase childhood vaccination. Low demand could reflect decision-making errors or binding constraints. But adoption has often remained low when such frictions are targeted (P. Dupas and E. Miguel, 2017). The low estimates may be driven in part by cultural factors. VSL is sensitive to religious views, with evidence showing the parameter is substantially lower in samples with more religious and fatalistic beliefs (Colmer, 2020; Lima, 2020). Kenya is highly religious, with one of the highest rates of belief in an afterlife globally (J. Evans et al., 2025).

The second component of this study validates that beliefs are measured without non-classical error. Two leading threats are experimenter demand effects and systematic misreporting. Since the concern is a gap between reported and decision-relevant subjective beliefs (I allow for arbitrary deviations from the 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. Using any combination of arms to instrument for beliefs yields similar estimates. Helmet demand also falls in response to the low signal, providing revealed preference evidence of belief updating.

Another design feature that allows for tests of misreported beliefs draws on the logic

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

[3] 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.

of local average treatment effects. Instrumenting for posteriors using treatment assignment weights observations linearly in priors, but one may construct instruments that weight observations by priors squared, yielding similar results. The stability of estimates suggests findings are not substantially affected by misreporting correlated with true beliefs, which includes most cases where agents systematically under or over-report probabilities (e.g. rounding or difficulty with 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 about risk, and instrumenting for beliefs using an observational shifter (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, Fischer, and Guiteras, 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, making them mechanically more likely to die on one, do not. Respondents also reported 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 statistically indistinguishable from the lower range of estimates from LMICs (Kremer, Leino, et al., 2011), but they are much smaller than the values used in 5 recent benefit-cost analyses of Kenyan programs. 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 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 reductions versus consumption. For many 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 large risks would lower their marginal utility of consumption, changing demand for safety when risks are non-marginal. Finally, many health interventions yield additional benefits, such as lower exposure to non-fatal morbidities or positive externalities, that are not captured by VSL.

This paper is among the first to estimate VSL using subjective beliefs and with methods that are robust to omitted variable bias.[4] This addresses two 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 B. K. Jack, 2015).[5] 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 value other non-market amenities. Goods such as time and environmental quality are economically important but difficult to value (Campbell and Brown, 2003; Greenstone and B. K. Jack, 2015). Policy decisions often rely on estimates that are not incentivized (e.g. contingent valuation surveys) when there is insufficient 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 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 belief data even when authors note concerns with assuming rational expectations (León and Edward Miguel, 2017a).

2.2 Study design and context

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 safety

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

[5] Kremer, Leino, et al. (2011) and Berry, Fischer, and Guiteras (2020) study the VSL of rural children, and León and Edward Miguel (2017a) and Ito and Zhang (2020) examine relatively wealthy populations.

improvement. Third, empirical mortality risk varies across the sample 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. In Kenya, there are an estimated 2.4 million drivers providing taxi services, combining for 22 million trips per day.[6] Transportation is dangerous. Data from the National Transport and Safety Authority (NTSA) reports 1,722 motorcycle deaths 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 all Kenyans.[7]

Despite the high risks, helmet use among taxi passengers is rare. Bachani et al. (2017) measure passenger helmet use at 3%. Helmeted passengers typically borrowed one from the driver, which anecdotally became uncommon after the COVID-19 pandemic due to hygiene concerns, further reducing helmet use. Low helmet adoption suggests that demand may be low. However, safe helmets have only recently become available at an affordable price. The *FIA Foundation* began offering helmets in Kenya in 2021, about a year before the experiment, and a local producer began manufacturing helmets at 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.

A natural concern with this sample is that consumers with low VSLs may select into using motorcycles. However, this setting was chosen because most Kenyans use motorcycle taxis, so extensive margin selection is limited. They are used similarly to traditional taxis 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 often taken on buses. 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 estimated 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.[8] Selection may be limited because alternative modes of public transportation are dangerous, crowded, and uncomfortable.

Ridership volume at the intensive margin may be more selected, for instance if richer consumers used motorcycles more often. A strength of this setting is that ridership volume is observed, allowing one to re-weight estimates to limit any effect of ridership on sample composition. Section 2.5 shows that doing so inconsequentially affects results, empirically suggesting that ridership selection does not meaningfully affect VSL estimates in this context.

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

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

[8]There were 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 ($22 \approx 53 \cdot 0.5 \cdot \frac{7.5}{7} \cdot 0.8$).

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 were excluded for safety. Survey locations are plotted on a map of Nairobi in Appendix Figure B.1.

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. Those 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 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 lacked time.[9]

Demographic information presented in Table 2.1 shows that the study reached a broad sample, suggesting that the results capture an informative, albeit not representative, population. Income aligns with representative samples of Kenya. 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.[10] 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.

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 affected 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.

[9]Surveyors reported recruitment rates and the share screened out due to helmet access to the field manager. The field manager reported that under 1% of approached passengers were screened out due to helmet access.

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

The treatments presented different studies of helmet effectiveness. Respondents first received an estimate of their unhelmeted mortality risk over the 5 year lifespan of a helmet.[11] They were then presented with one of two studies. Those in a "low treatment" arm were presented the results of Liu et al. (2008) which estimates, via a meta-analysis, that helmets reduce one's risk 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, so only the study and results vary across arms. The sources of information and the fact that the studies were conducted outside of Kenya were disclosed.

Both studies of helmet efficacy are from peer reviewed publications. There is a consensus that motorcycle helmets are effective, but there is uncertainty about exactly how well they work, particularly across contexts. 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.[12]

Helmet valuations

The study measured demand for a helmet using a Becker, Degroot, and Marschak (1964) willingness to pay mechanism. Respondents were asked the smallest cash gift that they would prefer to a free helmet. Surveyors then revealed a randomly drawn payment between 5 and 600 Kenyan shillings (Ksh). If the payment amount was greater than or equal to the respondent's bid, they received the cash. Otherwise they were given a helmet. The study offered a choice of a helmet versus cash to ensure that cash on hand did not limit demand. The mechanism elicits willingness to pay, not accept, since respondents did not give up a helmet they already owned. This aligns with standard VSL applications, which measure equivalent variation for proposed policies rather than compensating variation for losses. While preferences over existing health goods may differ due to endowment effects, most benefit-cost analyses use VSL in settings that closely match this experimental structure.

The maximum draw was based on assessments by an NGO and the helmet manufacturer that most valuations would fall below KsH 600. 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

[11] 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 in Kenya the year before the study by 22 million (trips/day) $\times$ 365. I combined the per trip estimate with 5 year ridership, obtained by scaling trips/week.

[12] An ethics appendix (Appendix D) details steps taken by the study to ensure that no respondents were harmed.

of Ksh 580, 15% of weekly median wages. A limitation of the BDM 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 exercise, making it unlikely that information would be recalled. In piloting, surveyors verified that respondents could not state the value.

In practice, helmet valuations often exceeded the wholesale price.[13] Appendix Figure B.2 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 ($p < 0.1$, $p < .05$ if baseline risk is above 1 in 100,000), validating that valuations affect real outcomes. I also show 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.

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 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.

2.3 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 $i \in \{1, \dots, \infty\}$. 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 expectations, which I test empirically.

[13]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.

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 deviate 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 probability of an accident and the effectiveness of a helmet given an accident. Therefore, suppose that the agent has a prior about the per trip risk of suffering 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

$$r_{iu} = 1 - \int_{Pr(A=1|\mathcal{I}_i)=0}^{1} [1 - Pr(A = 1|\mathcal{I}_i)]^{n_i} dPr(A = 1|\mathcal{I}_i) \tag{2.1}$$

$[1 - Pr(A = 1|\mathcal{I}_i)]^{n_i}$ 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 2.1.

I assume that helmeted individuals involved in an accident that would otherwise be fatal are deterred from continuing to use motorcycles.[14] 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

$$r_{ih}(z_i) = \begin{cases} H_{i0} \cdot r_{iu}, & z_i = 0 \\ H_{i1} \cdot r_{iu}, & z_i = 1 \end{cases} \tag{2.2}$$

which is the unhelmeted risk times the perceived chance that a helmet fails to prevent death.[15] The subscript u refers to the agent's perceived unhelmeted risk and h to their helmeted risk.

Let p_i be the price of a helmet. The present value of utility from being alive is given by $u_a(x_i)$ where x_i is a vector of characteristics, such as income, health, and age. Denote flow utility of consumption by $u(c_i; x_i)$ and the expected utility from not being alive (which may be positive, e.g. if one believes in an afterlife) by $u_d(x_i)$. The agent's expected utility from purchasing a helmet is

$$U_{ih}(z_i) = \zeta_h + [1 - r_{ih}(z_i)] \cdot u_a(x_i) - p_i \cdot u'(c_i; x_i) + r_{ih}(z_i) \cdot u_d(x_i) + \epsilon_{ih} \tag{2.3}$$

ζ_h captures average utility from characteristics of helmets other than mortality risk reduction, such as injury protection, comfort, and cultural stigmas that may limit helmet usage in some settings. ϵ_{ih} denotes idiosyncratic variation in utility. Without purchasing a helmet, expected utility is

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

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

[14] 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.

[15] Agents may increase ridership if they perceive a helmet is more effective. Empirically accounting for a ridership response has little effect on estimates, so I omit this from the model for simplicity.

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

$$VSL_i \equiv \frac{dp_i}{d\Delta r_i} = \left(\frac{\partial U_{i,h-u}}{\partial \Delta r_i}\right) \Big/ \left(\frac{\partial U_{i,h-u}}{\partial p_i}\right) = \frac{u_a(x_i) - u_d(x_i)}{u'(c_i; x_i)} \tag{2.4}$$

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 VSL will be higher even if agents obtain no more utility from averting a death. Expected utility from not being alive may also vary substantially across contexts, especially with beliefs about an afterlife which may raise $u_d(x_i)$ substantially above 0.

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

$$\text{Full Information Rational Expectations:} \quad \Delta r_i = \Delta r_i^* \quad \forall i \tag{2.5a}$$

$$\text{Exogeneity and Logit Errors:} \quad \epsilon_{ih}, \epsilon_{iu} \sim_{iid} EV1 \tag{2.5b}$$

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

$$Pr(y_i = 1) = \Lambda\left(\zeta_h + \beta \cdot \Delta r_i^* - \alpha \cdot p_i\right) \tag{2.6}$$

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

Assumption 2.5a 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 they match in expectation (mean-zero differences would produce attenuation bias). This is plausible if consumers have the same information as researchers, but fails if they possess private information or lack access to the data used by researchers. Rational expectations has been assumed despite these documented limitations because of a lack of an incentive-compatible alternative (prior to this paper) capable of producing informative estimates.[17]

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

[17] For instance, León and Edward Miguel (2017a) 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."

Assumption 2.5b 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 require rational expectations (e.g. Ito and Zhang, 2020).

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 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

$$\text{Exclusion Restriction:} \quad Cov\left(z_i, \epsilon_i\right) = 0 \tag{2.7a}$$

$$\text{Instrument Relevance:} \quad Cov(\Delta r_i, z_i) \neq 0 \tag{2.7b}$$

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

$$\begin{aligned}
\epsilon_{iu} &= \zeta_h + \beta \cdot \Delta r_i(z_i) - \alpha \cdot v_i + \epsilon_{ih} \\
\alpha v_i &= \zeta_h + \beta \cdot \Delta r_i(z_i) + \epsilon_i \\
v_i &= \frac{1}{\alpha}\zeta_h + \frac{\beta}{\alpha} \cdot \Delta r_i(z_i) + \frac{1}{\alpha}\epsilon_i = \zeta'_h + VSL \cdot \Delta r_i(z_i) + \epsilon'_i
\end{aligned} \tag{2.8}$$

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 2.7a is equivalent to

$$\text{Classical Measurement Error:} \quad \mathbb{E}[\Delta r_i - \Delta r_i^t | z_i] = 0 \tag{2.7a$'$}$$

This holds since z_i only contains information about Δr_i and (by randomization) is independent of other drivers of v_i. Therefore 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.

2.4 Data and empirical specification

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 B.1 and the full instrument is available on the AEA RCT Registry.[18] The survey elicited 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). Respondents were then asked to list a more precise belief within the interval. This two-step approach reduced rounding in piloting.[19]

Belief elicitation was not incentivized because the object of interest is respondents' perceptions of their own mortality risk. Incentivizing accuracy would require benchmarking against some average risk, inducing misreporting if respondents believed their own risk differed from the mean or mistrusted the data (e.g. anecdotally many respondents had concerns politicians under-count deaths). However, helmet demand is increasing in beliefs about motorcycle risk and helmet effectiveness (Appendix Table B.2), indicating that measures do align with real decisions. Furthermore, the study design includes multiple features to test for demand effects, poor comprehension of small probabilities, or other misreporting that could bias VSL estimates as detailed in Section 2.5.

[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 2.1 summarizes sample demographics and balance across treatments. Appendix Table B.1 similarly shows motorcycle use and beliefs. Variables are generally balanced across arms, although there is some imbalance among the pure control (which had a smaller number of respondents).

Experimental VSL estimation

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

$$\begin{aligned} v_i &= \zeta_h' + VSL\Delta r_i + X_i'\gamma_0 + \gamma_1 r_{0,i} + \epsilon_i' \\ \Delta r_i &= Z_i'\pi + X_i'\pi_c + \pi_r r_{0,i} + \nu_i \end{aligned} \tag{2.10}$$

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 a measure of 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, 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. Abadie, Gu, and Shen (2023) shows that similar instruments improve asymptotic mean squared error.[20] Controlling for $r_{0,i}$ isolates the exogenous portion of $r_{0,i} \cdot T_i$ to make the instrument valid, but there is a risk that measurement error in $r_{0,i}$ could enter first stage predictions of Δr_i if the two measures were elicited the same way, attenuating estimates. I therefore use a measure of $r_{0,i}$ – the respondent's belief about the number of deaths per 10,000 motorcycle users over 5 years – that was measured using a distinct method from Δr_i.

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] The primary tables report homoskedastic standard errors and weak instrument robust confidence sets. I also report heteroskedastic robust weak IV confidence sets for the primary VSL estimates. Results are similar using GMM with heteroskedastic robust errors. Appendix B.2 provides details, including justifications for these decisions.

[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.

Alternative VSL estimation procedures for comparison

I compare the experimental VSL estimates to two traditional approaches used to estimate the parameter. First, I estimate

$$v_i = \zeta_h + VSL\Delta r_i^* + X_i'\gamma_0 + \epsilon_i \tag{2.11}$$

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 Edward Miguel (2017a) 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 Edward Miguel (2017a), I do not instrument for Δr_i^*.

Second, I report the estimates

$$VSL_i = \frac{v_i}{\Delta r_i^*} \tag{2.12}$$

This assumes rational expectations and that agents do not receive utility from any characteristics of helmets other than safety. This specification follows Berry, Fischer, and Guiteras (2020) and is similar to Kremer, Leino, et al. (2011). The method is typically used with health products, where assumptions are 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 Edward Miguel, 2017a; Kremer, Leino, et al., 2011).

2.5 Results

Estimates of the value of a statistical life

Elicited beliefs about the probability of a fatal motorcycle accident fall in plausible ranges and are consistent 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 1/4,000. Beliefs about deaths per 10,000 passengers and one's own risk are highly correlated ($R^2 > 0.1$), though agents perceive lower personal versus average mortality risk.

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

[22]Results are robust to excluding implausible responses or winsorizing beliefs (Appendix Table B.3).

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

Respondents did update beliefs about the effectiveness of helmets in response to the treatments (Table 2.2, columns 3-4), which provides the primary variation used by the interacted instrument. At baseline, agents *overestimated* the effectiveness of helmets: the mean control group belief was 79%, with many respondents reporting values above 90%. Exposure to the 70% ("high") and 42% ("low") effectiveness estimates led respondents to revise beliefs downward to 75% and 65%, respectively, with the difference between treatment arms significant at the 1% level.[23]

The change in beliefs about helmet effectiveness was large enough to shift respondents' perceived probability that a helmet would save their life (the product of accident risk and helmet efficacy). The dependent variable in Columns 5-6 of Table 2.2 equals 10,000 times this perceived probability. Relative to the control group, the low-treatment group reports values that are about 80 units lower on average (36%, $p < 0.01$). The high-treatment group does not differ significantly from control (the signal was close to priors) though point estimates move in the predicted 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 2.3. The preferred estimate in column 1 is \$523, suggesting low demand for mortality risk reduction. A weak-instrument robust 95% confidence set excludes values below \$214 and above \$1,043. Estimates are consistent across specifications, both with and without controls, and across instrument sets.

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). This instrument is speculative as victim exposure could be endogenous. However, the identifying logic is that accident realizations are driven largely by luck, given that respondents travel in similar areas. The aim of this exercise is not to replace information treatments. Rather, the value is that this variation cannot be driven by demand effects, helping check for this confound. Estimates are remarkably similar to those obtained from random variation, with the mean VSL estimate in column 4 at \$560.

[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.

The estimates pass a battery of robustness checks. Columns 1-3 of panel b report interacted IV estimates excluding one of the experimental arms, which tests if results are stable shifting beliefs to different parts of the distribution. These results control for fixed effects for the taxi terminal where the trip originated to improve first stage power to overcome the reduction in sample size. The estimated VSLs are similar to the primary estimates, although due to a weak first stage values below $5,160 cannot be rejected when the control arm is excluded. Column 4 reports treatment only IV results dropping the control, a pre-specified result designed to rule out any behavioral effects of receiving information that helmets are effective. The result is statistically indistinguishable from the treatment only estimates in panel a.

Appendix Table B.3 shows that results hold if beliefs about helmet efficacy are winsorized. Appendix Table B.4 shows that the results are robust to changes in planned future ridership (Panel a), subjective beliefs about helmet lifespan (Panel b), and are not driven by selection into high ridership (Panel c).[24] Appendix Table B.5 shows that IV probit results regressing receipt of a helmet on price (BDM draw) and perceived risk reduction (instrumented by treatment) also yields similar results, validating that results are not driven by valuations above the maximum BDM draw.

As detailed in section 2.5, the local average VSL identified weights individuals with higher priors about the riskiness of motorcycles more heavily. Appendix Table B.6 examines whether demographic covariates are correlated with this variable to test if the experiment is identifying the parameter over a particularly selected population. The only covariate significantly correlated with priors is gender, which I show below is not a significant dimension of VSL heterogeneity.

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. Some differences may be due to methodology, but this helps to confirm the importance of income.

Figure 2.1 plots estimates from all LMIC revealed preference estimates of VSL that I am aware of. I also include Greenberg et al. (2021), which examines US soldiers, for a high-income country comparison. Estimates from this paper are lower than most prior estimates, but similar to those from populations with comparable incomes. In fact, the only other revealed-preference VSL estimate from Kenya, Kremer, Leino, et al. (2011), falls within the 95% confidence set of the preferred estimate from this paper (Table 2.3, column 1). Berry, Fischer, and Guiteras (2020) estimate median VSLs near $0 and just over $4,000 in Ghana (corresponding to two different treatment effect estimates). Ito and Zhang (2020) and León and Edward Miguel (2017a) find much higher VSLs among wealthier populations within

[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 -$54 to 1,438).

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 Edward Miguel (2017a) implies a coefficient of relative risk aversion of $\theta = 2.6$.[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 rises sharply with income.

Welfare implications of the estimated VSL

The welfare implications of these VSL estimates depends on the values that are currently used in benefit-cost analysis. If different VSL values are considered, then using these estimates could better align policy with preferences. A strong assumption underlying this welfare exercise is that the VSL identified is informative for the populations affected by the policies considered. This assumption is standard in practical applications of VSL: benefit–cost analyses either adjust rich-country VSLs for income differences or apply LMIC estimates to populations with similar income levels. I take a slightly more conservative approach by focusing on policies affecting Kenyans.

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.[26] I re-calculate benefit-cost ratios (BCRs) using the preferred VSL estimate and compare them to the original findings. Appendix Figure B.3 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,[27] 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

[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 $\theta \approx \log(700,000/523) \div \log(75,000/4,750) = 2.6$

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

[27] 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 2.4. 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.

I make two econometric changes to improve power. First, I pool high treatment and control observations, where posteriors are similar (Table 2.2), to limit small cell sizes within heterogeneous groups. Second, I use an interacted first stage with $r_{0,i}$ set to a respondent's prior about their own mortality risk. As discussed in section 2.4, this risks attenuating estimates if measurement error in $r_{0,i}$ is strongly correlated with that in Δr_i. Consistent with some attenuation, VSL estimates fall to $224 (about half the preferred estimate) with this estimator. I use this instrument despite this limitation because it greatly increases first stage power, making heterogeneity analysis feasible. Measurement error would most likely bias estimates towards 0, allowing one to draw inferences from the results, although they should be interpreted cautiously. Reassuringly, I detect statistically significant heterogeneity with respect to income and health, aligning with theoretical predictions.

I first examine heterogeneity by age (Table 2.4, column 1). I estimate similar VSLs below and above the median value. This result aligns with prior empirical work studying VSL over the life-cycle which finds that the parameter peaks near age 40 (Aldy and Viscusi, 2007).

Estimates of VSL heterogeneity by income show a strong positive relationship, matching theory. I consider heterogeneity with respect to asinh(wages) in Table 2.4, column 2, demeaned so that the first coefficient captures average VSL. [28] 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. This value is particularly striking because income is measured using a short question likely to attenuate the elasticity. Thus the estimate of 1.14 likely represents a lower bound, supporting the view that VSL increases sharply with income. VSL estimates are also about $500 higher among those with above-median income (Table 2.4, column 3) and reported health (Table 2.4, column 4).

In contrast, there is not significant heterogeneity with respect to having children, digit span recall (a cognitive test of memory), education, or gender. This suggests that the fact the study sampled more men than women is not likely to substantially influence estimates.

Are VSL estimates biased by non-classical measurement error?

The primary identifying assumption in this paper is that any gap between agents' true subjective beliefs and their reported beliefs reflects only classical measurement error. Research shows that agents decision-relevant beliefs can be shaped by heuristics like rounding (Lacetera, Pope, and Sydnor, 2012). This renders common tests of belief accuracy (such as bunching analysis) uninformative in this context, since the method allows for arbitrary differences between subjective beliefs and the empirical truth. To address this, the experi-

[28] 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).

mental design incorporates features that facilitate direct tests of belief validity. The results support the view that expectations are elicited with primarily mean-zero noise, suggesting that the VSL estimates are consistent.

I 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 arms included under classical error. However, if beliefs were misreported – unless misstatement was proportional across arms – estimates would differ. For example, if lower reported helmet efficacy in the low treatment arm reflected 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 2.3), 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 B.3 presents a formal argument, based on the literature on local average treatment effects (e.g. Abadie, Gu, and Shen, 2023; Mogstad, Santos, and Torgovitsky, 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. 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 produce different estimates, unless errors are orthogonal to true beliefs.

For instance, some agents report very high risks of riding motorcycles. 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.[29] 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.

Hawthorne effects are further ruled out by the reduction in willingness to 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 may plausibly generate demand effects. Lastly, as noted above, VSL estimates based on variation from accident victim exposure are statistically indistinguishable.

[29] Appendix Figure B.4 illustrates this in the case of rounding.

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 direct elicitation of small probabilities. I estimate a VSL of $164, 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.

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. However, measurement error could obscure a relationship. I therefore instrument for final beliefs using another measure (deaths/10,000 users), which addresses bias from independent classical errors (Gillen, Snowberg, and Yariv, 2019). This yields a statistically significant positive relationship, but the coefficient is below 0.001. I reject equality to 1 (the rational expectations benchmark) with 99.9% confidence. Because this test may be sensitive to correlated errors across belief measures, I supplement the analysis with two additional data sources to examine how agents form expectations.

First, the survey asked respondents to list the information sources they used to form beliefs, plotted in Figure 2.2. 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. I use trip volume and duration as empirical shifters, since more frequent motorcycle use mechanically increases risk, particularly conditional on taxi terminal fixed effects. In fact, 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 a victim.

If rational expectations held, one would expect positive coefficients on the empirical shifters of risk. Table 2.5 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 indicate a violation of rational expectations.

One remaining concern is that accident history may reflect risk heterogeneity observable to respondents but not to the econometrician, in which case rational expectations could hold under a more complete model of empirical risk. At a minimum, the results highlight the sensitivity of rational expectations to the researcher's data. To further probe expectation formation, I test whether reported five-year risks vary with rainfall on the survey day. Weather should not affect long-run accident risk, but rain heightens the salience of motorcycle dangers. Respondents reported roughly twice the risk if it rained, indicating a departure from rational expectations (Table 2.5, Column 5).

I next examine if typical approaches to estimate VSL, which assume rational expectations, consistently recover the parameter in Table 2.6.[30] Estimates from typical methods fall outside the experimental confidence sets, and different observational approaches yield starkly different results.

Columns 1 - 4 report estimates of VSL obtained from Equation 2.11. 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 2.12 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 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 significantly smaller than the experimental value. This suggests that both endogeneity and rational expectations violations may affect VSL estimates that rely on alternative methods in this setting.

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 dimensions of helmets other than their life-saving potential, particularly injury prevention. To limit this, the information treatments excluded injury content and no injury-related questions were asked. If injury beliefs did shift, the VSL estimates would be upper bounds so long as perceptions of helmets efficacy at preventing injury moved in the same direction as consumers' beliefs about their

[30] 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.

efficacy at preventing death, not affecting the conclusion that demand for reductions in mortality risk is low.

Second, VSL may be underestimated if agents did not plan to regularly use helmets. Piloting indicated that (consistent with the fact that helmets are costly) respondents believed that they would regularly use them, so such bias seems unlikely (although possible). If planned use were infrequent, measured helmet efficacy would overstate the true risk reduction, with the bias increasing in the level of reported effectiveness. VSL estimates would thus differ across arms, since the high treatment group would be more affected than the low treatment. The similarity of estimates excluding arms therefore provides support that infrequent helmet use is not driving the results.

Resale is also improbable because respondents all accepted a helmet with quality tags and the box removed, which undermines resale value but avoids hassle if worn. And even if respondents anticipated using a helmet 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 B.4 shows that estimates are similar when adjusting for anticipated duration of use.

Monotonicity violations could affect the interpretation of estimates as an average VSL. But 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 mismeasured helmet valuations bias results? Research shows that the BDM is accurate in similar stakes environments (Berry, Fischer, and Guiteras, 2020). Moreover, a level shift in valuations would not bias VSL estimates, which are identified from differences in willingness to pay across arms. 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 \$678 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.

2.6 Discussion

This section examines whether the VSL estimates in this paper are plausible and considers potential limitations of the estimates and with leveraging VSL to value health investments in poor settings. I also describe the conditions needed to apply the tools presented in this paper to other amenities.

Is an average VSL below annual income plausible?

The average VSL estimated in this study is 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. Section 2.3 highlights that the parameter is a function of one's expected utility from being alive minus perceptions about utility from not being alive. As a result, VSL may be sensitive to cultural and religious beliefs (Colmer, 2020). For instance, those with strong beliefs in an afterlife are likely to have VSLs that fall well below the lifetime utility of consumption because their expected utility from not being alive is high. Empirical evidence validates this view, finding that VSL falls sharply with strength of religious (Lima, 2020) and fatalistic views (León and Edward Miguel, 2017a). Fatalistic beliefs have been shown to limit HIV testing in Kenya (Oo et al., 2024), and the country has near universal belief in an afterlife (J. Evans et al., 2025).

Estimates also align with a consistent finding that demand for effective preventive health products is often low in LMICs (P. Dupas and E. 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 Pascaline Dupas (2010) show that only 40% of pregnant women in Kenya pay \$0.60 for a 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, Richardson, and Lengeler, 2018). Ashraf, Berry, and Shapiro (2010) find that demand for water treatment, which prevents a leading form of child mortality in diarrhea, drops 30 percentage points when prices rise from \$0.10 to \$0.25. Uptake of deworming pills falls by 60 percentage points when costs rise by \$0.30 (Kremer and Edward 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 child mortality by at least 0.5 percentage points in India (Kumar, 2024). Moreover, P. Dupas and E. Miguel (2017) note that demand for such products typically remains limited when capital or psychological constraints are mitigated, consistent with a low VSL.

Limitations of VSL in developing countries

In rich countries, welfare gains from prospective regulations or health interventions are often driven primarily by the value of lives saved. But a low VSL does not necessarily imply that investments in safety or public health institutions are not welfare enhancing in developing countries.

This conclusion would follow if preferences for mortality risk reduction and protection from non-fatal morbidity scaled proportionally with income. However, theory suggests the opposite. The same concavity of consumption utility that lowers willingness to pay for marginal reductions in mortality risk at low incomes may increase the relative disutility of non-fatal health shocks. Although speculative, this raises concerns about the common practice of valuing health interventions primarily through lives saved (with fixed adjustments for disability) in poor settings.

One can model health investment as a problem where agents receive concave utility from consumption each period and investments in safety increase survival odds (Hall and Jones, 2007). This framework implies that VSL rises nonlinearly with income: the marginal utility from higher consumption in a period falls as consumption rises, whereas expected utility is linear in survival probability. Non-fatal health shocks enter differently. Rather than increasing the probability that an agent survives, they reduce quality of life (consumption) conditional on survival. When consumption is low, such losses translate into large utility costs, implying that the relative disutility of morbidity to mortality risk should rise as income falls. Intuitively, the quality of life reduction from an illness or disability may be higher if one lacks resources to cope (Gertler and Gruber, 2002a).

This argument raises the possibility that poor consumers may have much higher demand for health interventions than their VSL suggests, since most programs that lower mortality risk also reduce morbidity risk. This conjecture is consistent with the fact that helmet demand often exceeds production costs in this setting despite the low VSL. However, the design of the experiment only allows me to determine that the intercept of agents' utility function, which captures all determinants of utility other than helmets' lifesaving potential, is positive. Future research designed to estimate demand for protection from non-fatal health shocks may therefore be valuable to inform whether VSL captures the main sources of agents' utility of health investments in developing countries.

Behavioral biases and VSL

This paper relaxes two strong assumptions made in revealed-preference VSL estimates (rational expectations and risk exogeneity) and, as argued above, the resulting estimates are consistent with theoretical predictions and related empirical evidence. However, estimates of VSL could be affected by behavioral biases that this paper is not designed to test.

One such force is present bias. Preventative health investments require upfront costs to protect against uncertain future risks, which could depress VSL estimates if individuals exhibit non-standard discounting. P. Dupas and E. Miguel (2017) argue that evidence of demand for preventative health (especially in voucher experiments) is often inconsistent with present bias. However, direct evidence studying whether such behavioral biases affect VSL estimates would be valuable to determine whether agents in LMICs do in fact have weak preferences for mortality risk reductions.

A strength of the method introduced in this paper is that it could be readily extended to study such mechanisms. Identifying observational settings where the variation needed to test if behavioral biases affect VSL is difficult. By contrast, one could combine this paper's framework with experimental tests for behavioral biases. This study therefore provides a first step towards testing further assumptions underpinning VSL estimates.

Limitations of VSL for valuing global health programs

These results should not be interpreted as a call to cut health funding to LMICs because the use of the estimates is not theoretically supported to value all health programs. VSL is relevant for decisions where a population faces a mortality risk-consumption choice. This informs many important decisions such as those discussed earlier. 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 2.3, 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

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

W is positive if $p \cdot VSL_\ell \cdot \frac{\alpha_\ell}{\alpha_h} > c$. Valuing lives saved instead at VSL_ℓ would understate welfare gains by $\frac{\alpha_\ell}{\alpha_h}$, which may be orders of magnitude. Intuitively, low-income consumers have low VSLs due to a high opportunity cost of funds. If high-income consumers pay costs, low-income consumers face no consumption tradeoff so their marginal utility of consumption should not be used to price costs. This distinction matters because applications of VSL often value global health investments using local VSLs, mechanically reaching the conclusion that fewer resources should flow to poor countries (Ahuja et al., 2021; Agrawal, Sood, and Whaley, 2023).

Conditions needed to estimate demand for other non-market amenities

The method introduced in this paper – updating beliefs about the effect of a product on exposure to a non-market amenity and then eliciting demand for the product – may also be informative to study demand for other amenities. However, the applicability of this method may be constrained in some settings by the existence of a suitable product to evaluate.

Applying the method requires the existence of a product that clearly affects exposure to a non-market amenity, that the product has not been ubiquitously adopted, and that there is credible information about the relationship between the product and amenity that can be used to update beliefs. Such a set of conditions may be challenging to identify in high-income settings where product diffusion is rapid, although strong cultural forces may produce informative populations in some cases (e.g. bicycle helmet usage is rare in the Netherlands). In developing countries, the slow diffusion of products creates far more opportunities to apply such tools. For instance, the use of air purifiers and water purifiers is often limited in sub-Saharan Africa, creating a feasible environment to study preferences for clean air and water. This method may therefore be particularly valuable to value non-market amenities in developing countries.

2.7 Conclusion

This study introduces a 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 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 Kenya. This is consistent with a common finding that uptake of preventative health products is low in LMICs. 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 may have 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 its benefits can exceed those of some health investments. Theory supports applying VSL only when a direct health-consumption tradeoff is faced, so these findings do not support reallocating health funds towards rich countries. Furthermore, VSL excludes welfare gains from protection from non-fatal health shocks, which may be particularly consequential in poor countries. Research studying the disutility from such shocks would be valuable to determine whether focusing welfare calculations on lives saved is appropriate in LMICs.

Finally, theory 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 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. Similarly, introducing an experimental method for producing the variation needed to identify VSL raises the possibility of extending the method to test whether forces such as psychological biases contribute to low estimates.

2.8 Tables and Figures

Figure 2.1: Revealed-preference VSL estimates across studies

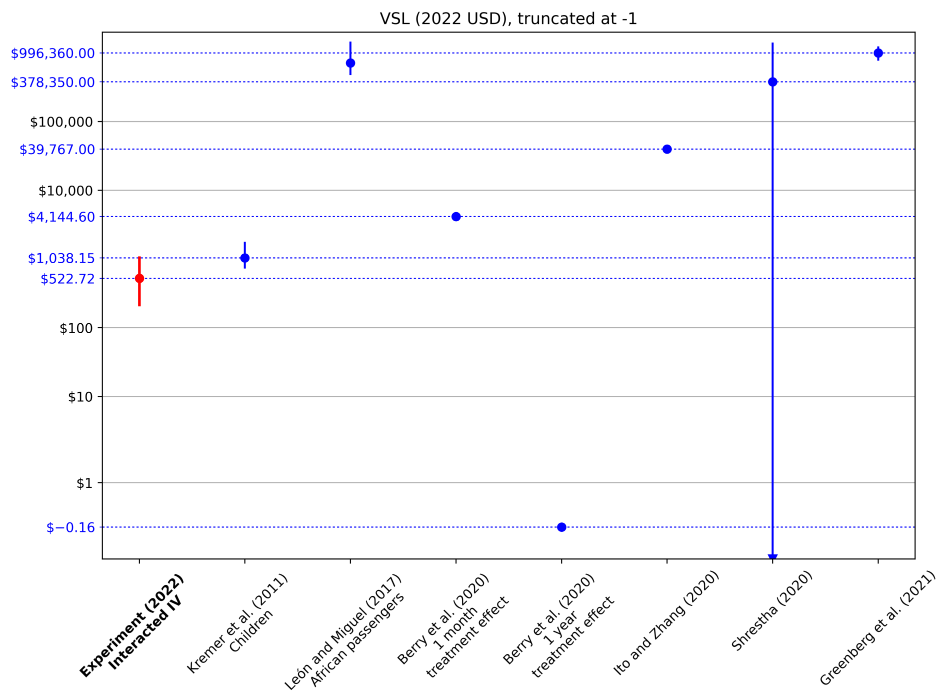


Figure 2.1 plots revealed-preference VSL estimates and (where available) 95% confidence intervals from this paper, Kremer, Leino, et al. (2011, rural Kenya), León and Edward Miguel (2017a, wealthy international travelers in Sierra Leone), Berry, Fischer, and Guiteras (2020, rural Ghana), Ito and Zhang (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 2.2: Information sources used to form beliefs about motorcycle safety

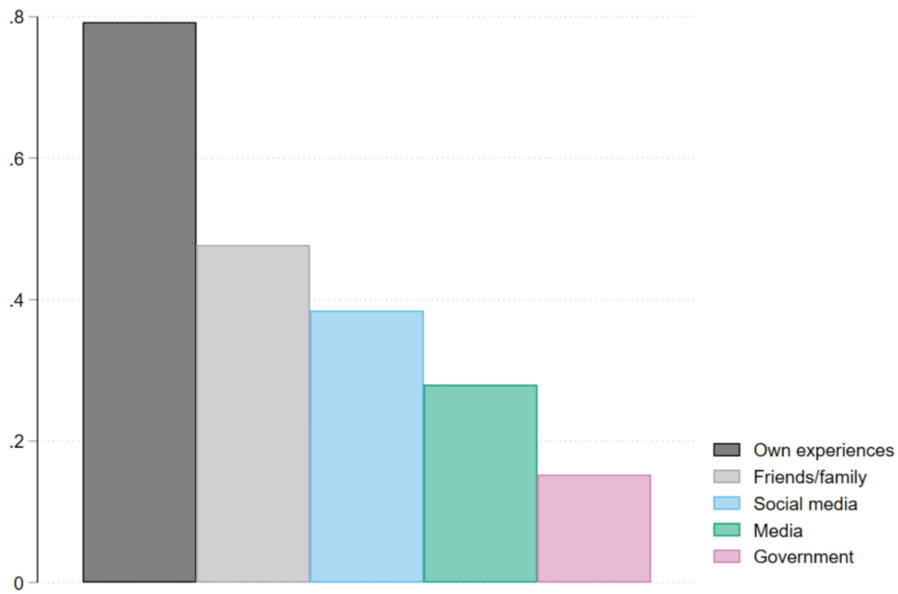


Figure 2.2 plots the information sources that respondents reported using to form beliefs about the mortality risk of motorcycles. Respondents were able to select multiple options, so the columns do not add to 1.

Table 2.1: 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 [8.126]	0.105 (1.159)	0.282 (0.572)	0.659 (0.589)	0.377 (0.590)
Female	0.352 [0.478]	-0.039 (0.060)	-0.064** (0.030)	-0.050 (0.030)	0.014 (0.029)
Health (1-5)	3.403 [0.667]	-0.034 (0.086)	-0.056 (0.043)	-0.012 (0.044)	0.044 (0.043)
Life expectancy	81.438 [6.947]	-0.134 (0.840)	0.457 (0.427)	-0.119 (0.439)	-0.577 (0.413)
Employed	0.898 [0.303]	-0.051 (0.043)	-0.041* (0.022)	-0.047** (0.022)	-0.006 (0.022)
Income ($1000s)	6.866 [8.427]	-0.150 (1.472)	0.635 (0.834)	1.017 (0.862)	0.384 (0.956)
$\mathbb{E}$[Wage in 5 years] /wage today	6.166 [11.685]	-0.636 (1.761)	-0.925 (0.710)	-1.422* (0.734)	-0.491 (0.667)
1(children)	0.774 [0.418]	0.033 (0.053)	-0.002 (0.027)	-0.014 (0.028)	-0.012 (0.027)
Digit span recall	3.020 [1.391]	-0.156 (0.183)	-0.006 (0.089)	-0.010 (0.091)	-0.002 (0.088)
Years of education	12.111 [2.869]	-0.468 (0.371)	0.161 (0.184)	-0.025 (0.189)	-0.186 (0.181)
Primary school complete	0.964 [0.186]	-0.035 (0.026)	0.007 (0.013)	-0.022 (0.013)	-0.029** (0.013)
Secondary school complete	0.722 [0.449]	-0.047 (0.058)	0.024 (0.029)	-0.008 (0.029)	-0.031 (0.028)
College degree	0.241 [0.428]	-0.054 (0.055)	-0.003 (0.028)	0.014 (0.028)	0.018 (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 2.2: 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 2.3: Value of a statistical life: Primary estimates

Panel A: Full sample	(1) Interacted	(2) Interacted	(3) Treatment only	(4) Know victim
VSL	522.72	503.16	347.26	559.99
	(182.76)	(182.72)	(259.49)	(225.82)
Cragg-Donald F-stat	12.33	12.34	10.99	16.24
Weak IV Robust Confidence Set	[213.95, 1,042.75]	[189.98, 1,016.72]	[-160.17, 1,066.93]	[193.40, 1,230.60]
Confidence Set: Robust SEs	[10.05, 1,286.50]	[23.60, 1,309.24]	[-105.89, 1,003.85]	NA
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: Estimates leaving 1 arm out	(1) High treatment Interacted	(2) Low treatment Interacted	(3) Control Interacted	Control Treatment only
VSL	383.39	576.99	787.15	721.11
	(179.71)	(273.64)	(542.92)	(453.95)
F-stat	17.00	9.92	3.89	10.46
Weak IV Robust Confidence Set	[42.70, 800.46]	[100.72, 1,405.38]	[-118.87, 5,160.10]	[-65.27, 2,225.04]
Inversion test	CLR	CLR	CLR	AR
Observations	963	905	982	982
Controls	BL Risk	BL Risk	BL Risk	BL Risk

Standard errors in parenthesis.

Panel A reports regressions using observations in all experimental arms. Panel B reports results leaving one experimental arm out. "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. All estimates control for baseline risk beliefs. Column (4) of panel A reports estimates that instrument for beliefs using exposure to a motorcycle accident victim, with robust standard errors in parenthesis. This instrument is speculative as knowing a victim may be endogenous, but is included to show that plausibly exogenous observational instruments align well with experimental estimates. Weak instrument robust confidence sets are reported in brackets, using the indicated tests. Weak IV and heteroskedastic robust confidence sets are also reported in panel A. Columns (1) - (3) of panel B include taxi terminal fixed effects. Excludes 35 observations where surveyors reported motorcycle taxi drivers pretending to be passengers.

Table 2.4: 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 2.5: 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 2.5 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, Chernozhukov, and Hansen, 2014a). 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 2.6: Value of a statistical life: Estimates using alternative methods

	Rational expectations					Subjective beliefs
	(1) OLS	(2) LASSO	(3) LASSO	(4) LASSO	(5) $\frac{WTP_i}{\Delta r_i^*}$	(6) Beliefs
Panel A: 42% helmet effectiveness treated as truth						
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						
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 eq. 2.11. Column (5) reports the mean of eq. 2.12. Panel A uses empirical helmet effectiveness from Liu et al. (2008). Panel B uses Ouellet and Kasantikul (2006). Column (6) uses beliefs about deaths/10,000 riders to instrument for agents' beliefs about the risk reduction of a helmet. 35 observations where taxi drivers pretended to be passengers are excluded.

Chapter 3

Can Cash Transfers Save Lives?

3.1 Introduction

The gradient between health and socioeconomic status — across societies and across individuals within societies — is one of the most widely documented correlations in the social sciences (Preston, 1975; Cutler, Lleras-Muney, and Vogl, 2012; Lleras-Muney, Schwandt, and Wherry, 2024). Studies often show a concave relationship (A. S. Deaton and Paxson, 2004; Cutler, Lleras-Muney, and Vogl, 2012), suggesting that poverty reduction in low-income settings may have particularly important implications for core health outcomes such as mortality. For instance, low- and middle-income countries (LMICs) still bear a disproportionate burden of child mortality (Burstein et al., 2019). However, causal evidence on the poverty-mortality relationship has been limited due to the need for both large-scale data collection (to achieve adequate statistical power) and credible exogenous variation in socioeconomic status.

The rise of unconditional cash transfer (UCT) programs provides an opportunity to study the causal effect of exogenous income gains on mortality. Over 100 LMICs have introduced UCT programs in the past two decades (Stedman, 2023), and a growing experimental literature has studied their effects on a wide range of development outcomes (Bastagli, Hagen-Zanker, and Sturge, 2016; Crosta et al., 2024). As UCTs become more established as an anti-poverty policy tool, interest has grown in understanding whether benefits accrue to the children of recipients, which could amplify the direct positive effects on recipients that have been documented in the short- and medium-run, improving UCTs' cost effectiveness. Yet despite the proliferation of studies on UCT programs (including many RCTs), it has remained challenging to experimentally examine impacts on a central marker of child well-being: whether they survive infancy and their first five years of life. Randomized evaluations of UCTs to date have typically lacked the large sample size and longitudinal data necessary to precisely estimate impacts on relatively rare but important outcomes such as child mortality.

This article presents one of the first experimental studies of the effects of unconditional cash transfers on child mortality. We evaluate the mortality effects of a program that distributed one-time UCTs, equal to 75% of average annual household expenditure, to over 10,000 households from randomly selected villages in western Kenya between 2015 and 2017.[1] This study leverages new census data recording over 100,000 births spanning more than a decade to examine the short- and medium-run effects of these UCTs on child survival.

The central empirical finding of this study, which was pre-specified, is that the cash transfer treatment led to a large and significant reduction in infant and child mortality during the cash transfer period. Results from the study's primary specification, which accounts for spillover effects within- and across-villages, indicate that infant mortality fell by over 19 deaths per thousand births among recipient households with a pregnancy during the transfer disbursal period. This represents a statistically significant 48% decline in infant mortality relative to the mean in control villages (p-value < 0.01, multiple-hypothesis testing (MHT) adjusted $p = .04$). We find similar results for under-five mortality (a reduction of 45%, MHT adjusted $p = .04$), as well as when we estimate a reduced-form OLS specification that builds directly on the two-stage research design, which randomizes treatment at both the village level and sublocation level (the econometric models are described in detail below). However, impacts are transitory: effects on mortality dissipate rapidly after the UCT period ends or, at the household level, when a household received cash before a female member became pregnant. This result aligns with prior evidence that households' marginal propensity to consume is high in this setting (Egger, Haushofer, Edward Miguel, Niehaus, et al., 2022b), suggesting that UCTs need to be delivered near the time of birth to yield mortality benefits in settings where saving is limited.

Several features of the research design and additional analyses indicate that the main result is unlikely to be confounded by factors such as measurement error or experimenter demand effects. The birth census — based on family recall of vital events — follows the same methodology used by prominent data sources such as the Demographic and Health Surveys (Romero Prieto, Verhulst, and Guillot, 2021). Evidence from the public health literature further supports the reliability of such data, as births and deaths are sufficiently salient to be accurately reported (Rao et al., 2003; Lyons-Amos and Stones, 2017; Nareeba et al., 2021). The large-scale, longitudinal design of the census also allows for extensive validity checks. We find no significant pre-treatment differences in infant or child mortality between treatment and control households during 2011–2014, or in households receiving cash just after a child turns 1 (or 5 respectively). In addition, the data are internally consistent across both cross-sectional and intertemporal dimensions: mortality varies as expected with household wealth, seasonality, and aggregate shocks. Finally, the estimates align closely with magnitudes implied by non-experimental variation, such as the cross-country income–mortality gradient, providing further reassurance about their validity.

The second part of the paper examines the mechanisms that might explain these reductions in child mortality. We first assess whether selection into fertility, an important

[1] Egger, Haushofer, Edward Miguel, Niehaus, et al. (2022b) document large economic gains among recipient and non-recipient households and local firms one and a half years on average after the transfers were implemented.

determinant of child health in other settings (e.g., Baird, McIntosh, and Özler, 2019), can account for effects. We document a modest, transient 10% increase in the share of women giving birth during the cash transfer period that dissipates thereafter and leaves completed lifetime fertility as well as total fertility over the entire post-UCT period unchanged.[2] Importantly, selective fertility choices cannot account for the decrease in mortality: the fertility effects appear mostly *after* the peak of the reduction in child mortality, which is concentrated among households whose eight-month window of transfer disbursal overlapped with an infant's birth month. Because households were informed of their eligibility within a month of their first payment, treatment status was unknown to these households at the time these pregnancies began, a clean test ruling out selection effects. Consistent with this interpretation, we find no significant change in fertility within nine months of a household's start date, nor any significant differences in mothers' observable characteristics.

Second, we examine changes in healthcare utilization. To do so, we bring in multiple rounds of household survey data (covering a representative sample of over 10,000 households), surveys of local health clinics and hospitals, and a dataset of drive times to health facilities obtained by equipping surveyors with GPS devices over the course of hundreds of trips throughout the study area. The most recent survey round (2024-25) collected particularly detailed information on antenatal, delivery, and postnatal healthcare utilization for births during the 2011-2023 period. During the census, we also completed verbal autopsies (VAs) for child deaths using the current World Health Organization methodology and assigned causes of death using a machine-learning classification algorithm (World Health Organization, 2022; Institute for Health Metrics and Evaluation, 2025). [3]

Several analyses indicate that access to and utilization of antenatal and delivery services played an important role in increasing child survival. First, analyzing mortality reductions by cause of death (as classified by verbal autopsies), we find declines across most major cause categories but show that the largest share of the overall effect is driven by birth complications and neonatal deaths, with mortality in the corresponding cause category falling by 75%. Second, we find that transfers are associated with a 45% increase in the rate of hospital deliveries; in this setting, hospitals are typically staffed by physicians and provide more extensive delivery services than local clinics but are far more expensive (Institute for Health Metrics and Evaluation, 2014). The increase in hospital deliveries occurred mainly among households with below-median travel time to such a facility. However, while mortality reductions are also somewhat larger in areas near hospitals, they remain large (and not statistically distinguishable) in regions of the study area far from hospitals where there is no significant evidence of greater healthcare use. Healthcare utilization alone, therefore, cannot

[2]This pattern aligns with prior evidence on limited fertility responses to income shocks in poor countries (Chatterjee and Vogl, 2018; Carneiro et al., 2021).

[3]In settings such as Kenya in which physical autopsies are rare and vital records incomplete, VA is considered the state-of-the-art method for determining causes of death at scale (Gacheri et al., 2014; Serina et al., 2015; Amek, Odhiambo, et al., 2014; Amek, Van Eijk, et al., 2018), and was validated in the study area by the Kenya Medical Research Institute (KEMRI), the country's flagship health research institution. KEMRI staff also trained this project's field staff in performing VAs.

fully explain the observed child mortality reductions.

Third, we turn to improved maternal health and nutrition as a plausible additional mechanism. Egger, Haushofer, Edward Miguel, Niehaus, et al. (2022b) document large overall improvements in nutrition and food security in recipient households for both adults and children. In this paper, we show that women around pregnancy particularly benefited in terms of improved health. Among female respondents in a household with a child aged one or younger, the transfers increase a pre-specified health index by 0.28 standard deviation units (p-value < 0.05) at the first endline survey. Compared to the control group, these women experienced 66% fewer major health problems since baseline (p-value < 0.05). In contrast, there is no significant impact on the health index among male respondents with infants in the household, nor among female respondents without infants in the household.[4] Several additional results support the view that increases in maternal health likely played a role. First, as noted above, the VA data shows that the overall reductions in mortality are driven overwhelmingly by a decrease in maternal and neonatal causes. In particular, reductions in stillbirths and pre-term delivery drive a meaningful share of the fall in mortality, and these are strongly linked to maternal health (László et al., 2013; Premji, 2014; Buffa et al., 2018). Second, the estimated mortality reductions are almost twice as large among older mothers, for whom birth complications are more prevalent, although differences by age are not significant.

Last, we examine maternal leisure, which may have reduced physical strain and contributed to improved maternal health during pregnancy. We estimate a large and significant 24 hour per week reduction (p-value < 0.05) in female labor supply among recipient households with a woman in the third trimester of pregnancy or the first three months following birth. This sharp decline – a 53% decrease compared to the control mean – is neither observed among men nor among women outside these months. This suggests that cash transfers may have enabled women at a critical period in pregnancy to reallocate work hours to rest (or other non-work activities). Kenya has one of the highest female labor force participation rates in the world, with women often performing strenuous physical tasks even during the late stages of pregnancy (Izugbara and Ngilangwa, 2010; Riang'a, Nangulu, and Broerse, 2018; Scorgie et al., 2023; International Labour Organization, 2025). Leisure is thought to be a critical input in the production of child survival (G. Miller and Urdinola, 2010), which is consistent with the literature on the positive impacts of parental leave documented in both rich and poor countries (Ruhm, 2000; Tanaka, 2005; Rossin, 2011; Nandi et al., 2016; Bartel et al., 2023).

This study provides some of the first *causal* evidence that income affects child mortality. Previous non-experimental studies of UCTs have documented a relationship (Richterman et al., 2023), raising the question of whether distributing cash is sufficient on its own to reduce child mortality or if the observed correlation is being driven by other contemporaneous health investments (Blattman and Niehaus, 2014; D. Evans and Kosec, 2016; Stedman, 2023).

[4]The absence of significant impacts on self-reported health among the full sample of adult recipients was reported by Egger, Haushofer, Edward Miguel, Niehaus, et al., 2022b.

Furthermore, McIntosh and Zeitlin, 2024 report the results of a randomized evaluation of a UCT program in Rwanda suggesting a 70% decline in child mortality among a sample of children that were already born when transfers were distributed, although the result is not significant once corrections for multiple hypothesis testing are applied (and is based on a relatively small sample, with 13 deaths in the control and 2 in the treatment group).

The expansive data analyzed in this study also allows us to understand the timing of and mechanisms underlying mortality effects, which may have policy implications. First, we find that effects are transitory, highlighting the importance of appropriately timing aid to yield infant mortality reductions. Second, evidence suggests that receipt of cash during pregnancy is important to reduce maternal physical strain, indicating that schemes that deliver cash to women only at or after birth may be sub-optimal. Third, we examine whether UCTs are effective as a substitute or complement to health infrastructure investments. We find some evidence that effects may be larger when health infrastructure is available, consistent with evidence on the importance of health infrastructure quality (Okeke, 2023). Yet effects are large and significant even far from such resources, where survey evidence shows little evidence of increased utilization. This suggests that UCTs may also be an effective tool to increase child health even in settings where health infrastructure is relatively limited. Further, a back-of-the-envelope calculation of the cost per child death averted suggests that the UCTs in this study, if targeted to women in the third trimester of pregnancy, are comparably cost effective to a number of WHO-recommended maternal and child health interventions even without taking into account the many other possible benefits of UCTs, i.e., on consumption.

This study also contributes to the literature on the timing of cash transfers and savings in low- and middle-income countries. Research shows that poor households often struggle to save without commitment devices (Ashraf, Karlan, and Yin, 2006) and that improving savings can enhance health investments (Pascaline Dupas and Robinson, 2013). These features have influenced the design of UCT programs such as the one studied in this paper, which often distribute cash in large lump sums to facilitate costly investments. In line with such savings barriers mattering for UCT efficacy, recipients often prefer lump sum transfers and sometimes prefer to delay them, suggesting attempts to align them with investment needs (Kansikas, Mani, and Niehaus, 2025). Yet little direct evidence exists about whether aligning UCT timing with actual investment opportunities consequentially affects program impacts, in part because it is challenging to identify exactly when lump sums are most needed. Our study setting of investment in child health therefore offers a useful opportunity since the time at which investments are needed, around the birth of a child, is well-defined. We document that cash receipt aligned precisely with critical stages of pregnancy leads to large reductions in infant mortality, but the effect vanishes if a household's final cash transfer was received before the third trimester of pregnancy. This finding suggests that the timing of transfers is important to their efficacy and that the high marginal propensity to consume observed for UCTs in low-income settings — which contributes to large transfer multipliers (Egger, Haushofer, Edward Miguel, Niehaus, et al., 2022b) — may, at the same time, limit child survival gains.

Finally, the results indicate that many households face a trade-off between using cash transfers to increase monetary consumption versus the consumption of non-monetary amenities, such as rest time, that have consequential effects on health but are infrequently captured in surveys. Economists have long recognized that GDP and poverty metrics omit consumption of non-monetary amenities such as leisure (Jones and Klenow, 2016; A. Deaton, 2016), and that such goods may be especially consequential for health in developing countries (Gertler and Gruber, 2002b). A small applied microeconomic literature also shows that the value of non-work time can represent a large share of the hourly wage in LMIC settings (Agness et al., 2025). Yet the extent to which such omissions change core results drawn from consumption measures is not well documented. We provide evidence that pregnant women increase consumption of non-work time (leisure) in response to UCT income gains, likely contributing to the substantial inter-generational health benefits that we document. Moreover, improvements in child survival are found precisely among those households with the smallest predicted monetary gains from transfers (over 1.5 years on average), suggesting a tradeoff between investments in maternal and child health versus future household income-generating activities. This pattern suggests that targeting UCTs solely based on their monetary variables (i.e., consumption, income or assets), as is common in practice (e.g., Haushofer et al., 2025), may lead to significant misallocation by neglecting the welfare gains from non-monetary investments that are consequential to recipient health and long-term inter-generational outcomes including child survival.

3.2 Intervention and Experimental Design

Context and Intervention

As part of the Kenya General Equilibrium Study (KGES), the NGO GiveDirectly (GD) provided unconditional cash transfers (UCTs) to poor households in rural Kenya, targeting households living in homes with thatched roofs as a simple proxy means-test for poverty. In treatment villages, GD enrolled all households meeting its thatched-roof eligibility criteria ("eligible" households); slightly more than one third of all households were eligible. Data from our surveys indicate that they are indeed poorer on average than other local households ("ineligible households"). For instance, assets per capita among eligible households at baseline were 64% lower than ineligible households.

Eligible households enrolled in GiveDirectly's program received a one-time series of three transfers totaling USD 1,000 (1,871 2015 USD PPP) via the mobile money system M-PESA, where the three tranches were disbursed over the course of 8 months: USD 80 a few weeks after enrollment and two equal sized tranches of USD 460 after two and eight months respectively. This is a one-time program and it was explained that no additional financial assistance would be provided to these households after their final transfer (and in fact none was provided). In total, the transfers constituted a shock of about 75% of household expenditure for eligible households, and of 15% of annual GDP in treated villages at the time they were distributed.[5]

Villages were phased into treatment starting in late 2014 and throughout 2015. The bulk of the payments were sent out during 2015 and 2016. The years 2015-2017 are thus most relevant for understanding the short-term impacts of UCTs on child mortality as many women who received cash transfers while pregnant during 2016 gave birth in 2017. Births from subsequent years allow us to test whether the effects of UCTs are persistent.

Background information on the Kenyan health care setting is also useful for understanding the analysis below. Health facilities in Kenya are classified into six levels: (1) community health units and community health workers; (2) primary care provided by dispensaries; (3) primary care provided by health centers; (4) sub-county hospitals (first referral); (5) county hospitals (second referral); and (6) national-level referral hospitals (N. Miller et al., 2024). Some level 2 facilities can perform deliveries, but they do not have inpatient care services. Level 3 clinics offer basic delivery services, though vary in whether or not they are staffed by physicians (only 15% of such facilities surveyed in this setting had physicians). Level 4 and 5 facilities are hospitals that typically have physicians (over 70% of the time based on survey data) and some specialists; there are no level 6 facilities in the study area.[6]

The Kenyan government has implemented several programs with the goal of reducing user fees for antenatal and delivery care.[7] However, implementation challenges and a lack of clarity around benefits means that in practice many women still pay substantial amounts out-of-pocket for these services (Orangi et al., 2021). In KGES surveys (described below), the total out-of-pocket cost of antenatal, delivery, and postnatal care averaged \$66 (USD 2023 PPP), with 10% of respondents spending over \$156. The average reported delivery cost alone in facilities staffed by a physician stood at over double (\$60) the cost of delivering in facilities without a physician (\$28), highlighting the trade-off faced by households. Additionally, even if some aspects of care are covered by government programs, qualitative work has documented patient concerns around referrals to higher level facilities due to the need to pay for transport (N. Miller et al., 2024). To illustrate the importance of travel costs in this setting (in which few households own cars or motorcycles), in 23% of control group births respondent mothers reported traveling by foot to a facility for delivery.

Experimental Design

Treatment assignment was randomized at two levels, the village level and the sublocation level. Within treatment villages, all households meeting GD's eligibility requirement received

[5] More details on program design and implementation are in Egger, Haushofer, Edward Miguel, Niehaus, et al. (2022b).

[6] A survey of health facilities targeted at level 3 locations in Siaya and bordering areas of surrounding counties conducted in the KGES project documented 94 level 3 facilities, 49 level 4 facilities, and just 13 level 5 facilities (and 94 level 1 and 2 facilities were also reached). Note that this implies that the median village has no health facility.

[7] In 2013, the Kenyan government introduced program to eliminate user fees, which became the "Linda Mama" program run by the National Health Insurance Fund in 2016. Other work has found that the program did not increase demand for maternal health services (Grépin, Habyarimana, and W. Jack, 2019).

the UCT. The second, sublocation level of randomization provided variation in local treatment intensity. Sublocations, an administrative unit directly above the village comprised of about ten villages on average, were randomly assigned to high or low saturation status: in high-saturation sublocations, two-thirds of villages were treated, while in low-saturation sublocations only one-third of villages were treated. This generated substantial spatial variation in local treatment intensity, which is used (as in Egger, Haushofer, Edward Miguel, Niehaus, et al., 2022b) to estimate spillover effects. In the analysis, we both directly follow the research design (in terms of village and sublocation assignment) and a spatial instrumental variables (IV) approach that utilizes all variation in cash transfer exposure over space, taking advantage of the idiosyncratic variation in local village assignments and the fact that villages could be located near other sublocations with different saturation assignment. Additionally, both treatment and control villages were randomly ordered for the program and for data collection visits, allowing us to assign an "experimental start date" to each village and explore effects related to transfer timing. These approaches are described below in Section 3.3.

Egger, Haushofer, Edward Miguel, Niehaus, et al., 2022b document that the UCTs led to significant increases in living standards for recipient households. Recipient households' marginal propensity to spend out of the transfer is approximately 0.8-0.9 over the first 1-2 years after the transfer, with the largest increase in spending concentrated over the first months, and the vast majority of spending occurring locally within the study area. At endline, an average of 1.5 years after the start of the program, they still report a 13% increase in consumption expenditures (including a 9% increase in food expenditures) and 26% increase in asset ownership, associated with substantial increases in food security, and the quality of the home environment (e.g. the quality of roofing materials). The increased spending by recipient households generated local (within 2 km) increases in economic activity, higher profits for firms, and higher wages, income and expenditure for non-recipient households. There is a positive but small 0.2% average increase in prices in areas that received more cash relative to those that received less. Taken together, Egger, Haushofer, Edward Miguel, Niehaus, et al., 2022b estimate a real transfer multiplier of around 2.5 from the UCTs. These substantial short-term local economic gains from a large cash transfer raise the possibility of child health effects from the program.

3.3 Data and Estimation

Data

The primary data for this paper was collected as part of a third endline round (EL3) of data collection for the KGES. This built on the project's baseline household censuses and surveys (2014-15), endline 1 household surveys (2016-17), and endline 2 household censuses and surveys (2019-22). The infant and child mortality analysis primarily uses EL3 household census data, which included birth histories for adult female household members; we augment the census data with household surveys from a representative sample of eligible and ineligible households to get additional details on birth experiences and healthcare utilization. We discuss each of these data sources below. We also make use of data on the location of health facilities and travel times, which we introduce when discussing mechanisms.

Endline 3 household census data

The endline 3 (EL3) household census took place from April to November 2023 in all study villages in Siaya, Kenya. The analysis focuses on households that were present at baseline (i.e., the start of the GiveDirectly program) and therefore have a clearly defined program treatment and eligiblity status. The household census captures all such households that still resided in the study region in 2023, which encompasses 94% of baseline households. Birth information for the approximately 6% of baseline households that had migrated out of the study area is captured in the representative household survey mentioned above, in which the study team attempted to track all sampled households wherever they had moved across Kenya (and succeeded in locating and surveying slightly over half of movers). These observations are then reweighted to maintain baseline sample representativeness.

The pre-analysis plan (PAP) specified two primary outcomes (Egger, Haushofer, Edward Miguel, and M. Walker, 2023): infant (under-one) mortality and child (under-five) mortality. For each measure, we focus on children born at least one or five years before the time of data collection, respectively, in order to have a consistent population for both the numerator (children who are deceased) and the denominator (children who are deceased plus children who have survived until the age of one or five). We thus examine effects on under-five mortality among children born through the end of 2017 (i.e., in the pre-program period and in the treatment period of 2015-17), while for infant mortality we can estimate impacts among children born up to 2021, which also includes the post-cash transfer period of 2018-2021.

In addition to collecting information on fertility and mortality, we also seek to assign causes of death for child mortality instances via verbal autopsies (VA) using the standard World Health Organization 2022 VA Questionnaire (World Health Organization, 2022). The PAP specified that we would examine whether there were cash transfer treatment effects on the main causes of death (Egger, Haushofer, Edward Miguel, and M. Walker, 2023). Verbal autopsy is considered the state-of-the-art survey-based method for determining causes of death based on self-reported information (as physical autopsies are rarely performed in many LMICs and vital records are incomplete), with previous literature having validated the accuracy of VA methodology and the associated machine-learning classification algorithm, including in the Kenyan study region (Gacheri et al., 2014; Serina et al., 2015; Amek, Odhiambo, et al., 2014; Amek, Van Eijk, et al., 2018). Though the literature also notes the limitations of VA relative to the administrative data available in wealthy nations, it is the most accepted approach to determining the distribution of cause of death in populations such as the present one.[8]

[8]The VA module includes a categorization for stillbirths in approximately 15% of cases. As noted below,

Overall follow-up rates were high in the EL3 household census: over 92% of households in the 653 study villages completed the birth history and child mortality modules. The response rate in the census activity was also nearly identical and not statistically different across treatment and control villages. In total, across the birth census and representative surveys designed to collect information on migrant households, we collected information on 101,405 births. When restricting attention to births in eligible households contemporaneous with the disbursal of cash transfers (2015-17), the population under study is 6,347 births.

Descriptive statistics for the full census of births as well as subsets of interest, such as births to transfer-eligible households, are presented in Appendix Table C.1. Across all births in the census, the infant (under-one) mortality rate was 33.7 deaths/1,000 births and the child (under-five) mortality rate was 46.7 deaths/1,000 births. These recorded infant and child mortality rates are similar to those estimated in Kenya by the United Nations Inter-Agency Group for Child Mortality Estimation (2025).[9]

Endline 3 household survey data

Household surveys were conducted at baseline, endline 1, endline 2 and endline 3 with a representative sample of censused households. Specifically, at baseline the KGES team sought to survey eight (8) eligible households and four (4) ineligible households per village; in cases where initially-targeted households were not available at the time of survey, "replacement" households were surveyed instead. Endline 1 sought to survey all households initially selected for surveys, as well as replacement households see Egger, Haushofer, Edward Miguel, Niehaus, et al., 2022b, for details. Endlines 2 and 3 maintained this sampling frame of households present at baseline while adding in newly-identified households, as described in Egger, Killeen, et al., 2024; since the focus of this analysis is on households present at baseline, we exclude those who later moved into the area from the analysis. The EL3 household survey tracking rate was over 90% and balanced across treatment and control groups (Appendix Table C.2).

Household survey data from the KGES project provides two main benefits to the analysis. First, household surveys tracked individuals that moved outside of the study area, allowing the analysis to account for child births and survival among individuals present in the study area at the time of transfers but that moved away (and thus were not captured in the census). This accounts for 6% of households from the detailed survey sample. In total, enumerators were able to survey 148 moved-away eligible households that had had at least one birth during this time period. For these households, we conducted the same birth history as in the census (in addition to the other survey questions described below), and we include these

the results are also robust to including only live births, which is sometimes the sample considered in analyses of infant and child mortality.

[9] Note that births to households present at baseline – which comprise approximately 77% of total births and are as noted above the primary focus of the analysis – exhibit virtually identical rates of infant and child mortality (as well as a similar average maternal age at birth) to the census data as a whole.

observations with sampling weights to reflect their proportion of the censused population in the child mortality analysis.

Second, household surveys gathered more detailed data than the census, providing a means to study potential mechanisms. Endline 1 (2016-17) provides short-term data on household living standards, assets, labor supply, health, nutrition and food security. We highlight some results previously reported in Egger, Haushofer, Edward Miguel, Niehaus, et al., 2022b and conduct further analyses using these data. Additionally, the EL3 household survey data provides detailed information on health behaviors and access that could serve as channels (e.g. hospital delivery, and antenatal and postnatal care). We specified the outcomes we would focus on as mechanisms in a pre-analysis plan (Egger, Killeen, et al., 2024).

Estimation Framework

Primary Econometric Specifications

We first estimate the following reduced-form specification, focusing attention on births in eligible households that were present at the time of the cash transfers:

$$\begin{aligned} y_{imhvs} = {} & \alpha_1 Treat_v + \alpha_2 HighSat_s \\ & + \lambda_{t(i)} + \rho_{g(i)} + \lambda_{t(i)} \times \rho_{g(i)} + A_m + \delta M_i + \epsilon_{imhvs}, \end{aligned} \tag{3.1}$$

where y_{imhvs} is an outcome of interest (i.e., infant mortality) for a birth i in household h, located at baseline in village v and sublocation s. The variable $Treat_v$ is an indicator for residing in a treatment village at baseline and $HighSat_s$ is an indicator for being in a high-saturation sublocation. The specification includes child year of birth fixed effects, denoted by $\lambda_{t(i)}$, and child gender fixed effects $\rho_{g(i)}$. We control for maternal age through the inclusion of indicator variables represented by A_m, where m denotes one of five age groups (under 20, 20-25, 26-30, 31-35, or above 35).[10] Standard errors are clustered at the sublocation level. While most of the data is derived from the birth census of the study region, as noted above we also survey a representative sample of households that migrated from the region and re-weight those observations by inverse sampling probabilities.

The coefficient α_1 captures the effect of the transfers on eligible households in treatment villages (relative to control villages) from two sources: the direct effect of treatment and any within-village spillover effects. The coefficient α_2 estimates cross-village spillover effects based on the research design of high- versus low-saturation sublocations. This estimation of cross-village spillovers is relatively coarse as it does not utilize all experimental variation, implicitly assuming that all spillovers are contained within villages and sublocations. The sum of α_1 and α_2 denotes the total effect of the transfers from all three sources (direct effects, within-village and cross-village spillovers). This linear combination of coefficients captures

[10] M_i is a vector of indicators for a missing value of a covariate. Where a covariate is missing, we set the value equal to the covariate's mean.

the effect of the transfers on eligible households in treatment villages in high-intensity sublocations relative to eligible households in control villages in low-intensity sublocations.[11]

Equation (3.1) is a straightforward and intuitive benchmark yet it does not capture the full spatial dimension of spillovers. One-third of villages in low-saturation sublocations still received transfers, and there is additional variation stemming from by the idiosyncratic placement of treatment and control villages as well as sublocation boundaries. In the following pre-specified analysis, which is based on Egger, Haushofer, Edward Miguel, Niehaus, et al. (2022b) and builds on Edward Miguel and Kremer (2004), we make use of the full spatial variation induced by the experimental design. These regressions allow us to compare areas that, due to the randomization, received more cash relative to areas that received less. We utilize the following regression specification:

$$\begin{aligned} y_{imhvs} \;=\; & \beta_1 Amt_v + \sum_{r=2}^{\bar{R}} \beta_r Amt_{v,r}^{\neg v} + \gamma_1 ShareElig_v + \sum_{r=2}^{\bar{R}} \gamma_r ShareElig_{v,r}^{\neg v} \\ & + \lambda_{t(i)} + \rho_{g(i)} + \lambda_{t(i)} \times \rho_{g(i)} + A_m + \delta M_i + \epsilon_{imhvs} \end{aligned} \tag{3.2}$$

The key coefficients are β_1, which captures the effect within treatment villages from both direct receipt of the transfers and within-village spillovers (where cash transferred to the village is captured in Amt_v), and the β_r terms, which capture the effects of cash transfers in other villages (not v) at different bands of radius r ($Amt_{v,r}^{\neg v}$) away from village v. $ShareElig$ denotes the share of baseline households eligible for a transfer in an area. The other terms are as in Equation (3.1). As in Egger, Haushofer, Edward Miguel, Niehaus, et al. (2022b), the maximum radius ($\bar{R}$) is found by estimating models with varying radii, then selecting the model that minimizes the Schwarz Bayesian Information Criterion (BIC).[12] The Schwarz BIC algorithm indicates that infant and child mortality effects are locally concentrated within 2 km of cash transfer receipt in nearly all cases, as was also the case for most outcomes in Egger, Haushofer, Edward Miguel, Niehaus, et al., 2022b.

The amount of cash an area received is a function of the share of households that were eligible at baseline ($ShareElig$), which is endogenous, and the share of eligible households that were treated. We thus instrument for transfer amounts using the share of eligible households treated in an area and the share of eligible households treated interacted with the share of eligible households, which is a valid instrument since the estimates control for the proportion of households that were eligible in an area at baseline.[13] To account for spatial

[11]Among other main analyses, we pre-specified a version of Equation (3.1) that clustered standard errors at the village level and focused on α_1 while considering α_2 as non-primary. We adjust standard errors to cluster more conservatively at the sublocation level as our analytic focus expanded to include α_2.

[12]For computational reasons, once the BIC increases at a radius, we stop searching and select the minimizing value over the earlier radii searched.

[13]Egger, Haushofer, Edward Miguel, Niehaus, et al. (2022b) only uses the share of eligible households treated as an instrument. By logic similar to Abadie, Gu, and Shen (2023), the revised instrument vector is efficient under homogeneous treatment effects because it captures the true first stage. The original PAP specified the instruments from Egger, Haushofer, Edward Miguel, Niehaus, et al. (2022b), but we noted that

correlation, we calculate spatial heteroskedasticity- and autocorrelation-consistent (HAC) standard errors using a positive definite kernel up to 10 km (Conley, 2008).[14]

The analysis focuses on the "average total effect" of the cash transfers on births in recipient households in high-saturation sublocations, which is defined as:

$$\begin{aligned}\widehat{\Delta y^r} &= \hat{\beta}_1 \cdot \left(\overline{Amt_v} | i \text{ born in recipient household in high-saturation sublocation}\right) \\ &+ \sum_{r=2}^{\bar{R}} \hat{\beta}_r \cdot \left(\overline{Amt}_{v,r}^{\neg v} | i \text{ born in recipient household in high-saturation sublocation}\right)\end{aligned} \tag{3.3}$$

If there are no effects of cash outside of the radius $\bar{R}$ (e.g., due to ambient effects in the study area), this equation estimates the average effect of the cash transfers on recipient households in high-saturation sublocations compared to a counterfactual in which no cash was distributed. Note that, to the extent that ambient spillover effects have the same sign as the direct effect over all radii, even beyond those chosen for inclusion in Equation (3.2), then this quantity is a lower bound on the true effect of the program.

When working with the full census sample of births to estimate overall effects, we mainly focus on Equation 3.2 since it captures all variation in cash exposure. However, this estimator relies on asymptotic consistency which, combined with the spatial dependence of the data, may yield biased estimates in finite samples. We therefore focus on reduced-form estimates from Equation 3.1, which have better finite sample properties, when considering heterogeneity estimates that effectively divide the sample into much smaller subgroups. We present equivalent heterogeneity results for census outcomes using Equation 3.2 in the Appendix, but omit similar heterogeneity results for the household survey data (given that the effective subgroup sample sizes fall considerably).

Secondary Econometric Specification

Date-of-birth data from the household census enables us to examine whether the effect of cash varies depending on the timing of receipt. We first restrict the sample to births where the transfer period — the experimental start date plus 8 months, since the three transfers were distributed over that time frame — intersected with an age group G. The relevant age groups G are defined as: (i) 9 months to three years before birth (treatment pre-pregnancy), (ii) within 9 months before birth up to the month before birth (treatment in-utero), (iii) the birth month and up to 28 days after birth (treatment at birth and as a neonate), (iv) 28 days after birth to one year after birth (treatment as an infant), and (v) one to three years after

this was an active research area and we might consider estimates based on recent advances. An amendment filed before the data was analyzed (but after data collection began) documented our intent to switch to the specification presented here. Standard errors are somewhat larger with the original approach.

[14]We use the kernel $K_{ij} = 1(d_{ij} < 10) \cdot \left(1 - \frac{d_{ij}}{10}\right)^2$ where d_{ij} is the Euclidean distance in km between i and j. We use this kernel, rather than a uniform kernel as in Egger, Haushofer, Edward Miguel, Niehaus, et al. (2022b), to ensure that standard errors are not complex-valued.

birth (treatment as a young child). For instance, if group G is in-utero, then observation i is included in estimates of the effects of receiving cash in-utero if and only if the intersection of the transfer period and individual i's birth date minus 9 months is not null.

This allows for the estimation of treatment effects among eligible households in treatment and control villages in group G as follows (restricting the sample to households in group G):

$$\begin{aligned} y_{imhvs} &= \alpha_{1,G} Treat_v + \alpha_{2,G} HighSat_s \\ &\quad + \lambda_{t(i)} + \rho_{g(i)} + \lambda_{t(i)} \times \rho_{g(i)} + A_m + \delta M_i + \epsilon_{imhvs} \end{aligned} \tag{3.4}$$

We focus on the treatment effect in high-saturation sublocations for each age group, namely, $\alpha_{1,G} + \alpha_{2,G}$, since some cells end up with modest sample sizes, favoring the better finite sample properties of this reduced-form specification (similar to the issues around estimating heterogeneity across subgroups noted above). In additional analysis, we consider the spatial IV estimator of dynamic effects by similarly restricting the observations into birth timing groups, then estimating equations (3.2) and (3.3) separately for each group; both approaches produce similar results.

Intuitively, this specification examines whether there are different effects on children that were exposed to cash at different times either before or after birth. Any effects of pre-birth exposure are more likely to work through effects on mothers (i.e., in terms of their nutrition, health, stress, and medical care received). This perspective helps to motivate the mechanisms examined below through which cash could affect child mortality, and the non-uniform sized age groups are motivated by key stages of pregnancy and early life.

Because the transfers were distributed over 8 months, many births fall into multiple G groups and so a single birth observation can be included in different estimates. Therefore estimates of the effects of cash when received in-utero may also partially reflect the effect of receiving cash in the birth month, and similarly for other birth timing groups. Estimates should therefore be interpreted as the joint effect of cash exposure in age group G and adjacent pregnancy stages since the design of the experiment does not identify "unbundled" coefficients without additional assumptions.[15]

[15]The original PAP specified a version of equation (3.4) that estimated the effect of cash received in each group on mortality jointly by calculating how much cash was transferred in each group G and then estimating one regression across the sample. While this could in principle separately identify the effects of cash at different stages, a recent literature in econometrics such as Callaway and Sant'Anna (2020) — which studies dynamic difference-in-differences but generalizes to this setting where observations that received cash in different stages serve as controls in joint estimation — shows that such methods can generate biased estimates. In fact, when we estimate the PAP specification, coefficient estimates suggest that cash *increased* deaths in each pregnancy stage, which is is soundly rejected by our overall estimates and consistent with consequential bias. We have opted to implement the more robust approach described here and interpret results accordingly.

3.4 Main Empirical Results

Graphical Analysis

The results are first presented graphically in Figure 3.1. Panel A plots simple means of infant mortality by year across 2011-21 for two central groups of interest within the population of eligible households: births in treatment villages located in high-saturation sublocations, and births in control villages located in low-saturation sublocations. The first group comprises censused children born in villages that received the highest average intensity of cash transfers, whereas the second group are those with the lowest average intensity of transfers.

A clear pattern is evident from Panel A: while high-saturation treatment and low-saturation control villages exhibit similar and not statistically distinguishable levels of infant mortality during the pre-period of 2011-14, a marked divergence occurs once cash transfers are distributed. Following the start of transfers, infant mortality in high-saturation treatment villages rapidly falls from 37.5 deaths per thousand births in 2014 to 20.5 deaths per thousand births in 2015, a 45% decline. By contrast, low-saturation control villages continue on their pre-COVID-19 trajectory of gradual improvement over time (absent a drought in 2017 which elevated infant mortality).[16] Once the disbursal of transfers ends, however, infant mortality in high-saturation treatment villages swiftly returns to the rates seen in low-intensity villages. Rates for both groups rise in this latter period, and particularly in 2020 and 2021, likely related to the effects of the COVID-19 pandemic.

Regression Analysis

Turning to the first regression results from the reduced-form specification, Panel B of Figure 3.1 displays estimates of the effect of living in a high-saturation treatment village (i.e., $\alpha_1+\alpha_2$ from Equation (3.1)) on infant mortality by year of birth. There are large impacts on infant survival during the period in which UCTs were disbursed. The coefficient estimates range from 15 to 24 fewer infant deaths per thousand births during the three years in the UCT period. Pooling the three years of the UCT period, the effect of transfers is statistically significant at the one percent level (p-value < 0.01, MHT adjusted $p = .04$), as shown in the regression analysis below. Though we lack the statistical power to detect differences in coefficients across years within the transfer disbursal period, the year with the largest treatment coefficient (2017) is also the year in which a drought affected the region, which is consistent with the view that UCTs reduce infant mortality most in settings with the greatest economic adversity, a point we return to below. The magnitude is similar and the statistical significance of the result remains unchanged (p-value < 0.01), however, even when the sample is restricted to the more economically favorable years of 2015 and 2016.

[16] Note that droughts have become increasingly frequent due to climate change (Nkiaka, Nawaz, and Lovett, 2022). Within East Africa, the frequency of droughts has doubled since 2005, from every six years to every three years (United Nations Office for Disaster Risk Reduction, 2023). This could jeopardize the long-standing trend toward reduced infant and child mortality in the region.

We do not detect meaningful treatment effects in any of the pre-period years, nor do we find significant persistent impacts in the post-UCT period.[17]

Table 3.1 reports this study's main results. Column 1 displays estimates of the reduced-form effect of UCTs on infant (under-one) mortality for eligible households during the period in which transfers were disbursed. Infant mortality declines by 17.9 deaths per thousand births in treatment villages located within high-intensity sublocations, a result statistically significant at the one percent level. This coefficient estimate represents a 44% decline in mortality relative to the low-saturation control mean of 40.2 deaths per thousand births.[18] The overall effect is driven both by the direct effect of cash transfer receipt including spillovers within treated villages (α_1), and the effect of cross-village spillovers (α_2); Egger, Haushofer, Edward Miguel, Niehaus, et al., 2022b had shown large effects both of direct cash transfer receipt and of local spillovers in terms of local economic outcomes and living standards. Similarly, when examining child (under-five) mortality in Column 2, there is a reduction of 17.6 deaths per thousand births during the UCT period, a 31% decline relative to the control mean, and once again both the direct effect and spillover estimates are large and negative. This pattern of findings also indicates that most of the reduction in under-5 deaths occurs in the first year of life.

Column 3 of Table 3.1 presents estimates on infant mortality from Equation (3.2), the instrumental variables specification which more fully captures the spatial dimension of spillover effects. From this specification, we find an average total effect (including direct, within-village and across-village spillovers) on infant mortality in recipient households of -19.5 deaths per 1000 births across 2015-17. This represents a 48% decline in infant mortality relative to the control mean and is statistically significant at the one percent level. We estimate significant effects both from transfers within one's own village (β_1) and from other nearby villages (β_2) at the 10% level. Column 4 reports spatial IV estimates of transfer impacts on child mortality, and estimates an average reduction of 25.6 child deaths per 1000 births during the UCT period, a 45% decline relative to control significant at the one percent level, and once again indicating that most of the reduction occurs in the first year of life. These last two effects remain statistically significant (at $p < 0.05$) when accounting for multiple hypothesis testing across the reduced-form and spatial IV results (using the Romano and Wolf (2005) step-down approach as pre-specified). In all, both specifications yield the same striking finding: cash transfers lead to declines of between 31% to 48% in infant and child mortality.

Once cash transfers cease, so do the impacts on infant mortality. In the post-UCT period, we do not estimate significant differences in infant survival between treatment and control

[17]Appendix Figure C.1 presents the main estimates for infant mortality aggregating across the main time periods, including the cash transfer period (2015-17) as well as the pre-period (2011-14) and the post-UCT period (2018-21). Appendix Table C.4 reports results separately for 2011-14, 2015-16, 2017, and 2018-21.

[18]Distinguishing between stillbirth and neonatal deaths soon after live birth can be challenging outside of medical settings. While we include stillbirths in the main analyses as child deaths, we present the main results excluding stillbirths (as determined by a verbal autopsy classification algorithm) in Appendix Table C.5 and find similar results in terms of magnitude and statistical significance.

villages, though the standard errors cannot rule out modest impacts in either direction.[19] We furthermore do not find significant changes over time in infant mortality among ineligible households who did not receive cash transfers. There are also no meaningful treatment effects on these non-recipient households overall, who as noted above were, on average, considerably richer at baseline in addition to having substantially lower infant and child mortality rates (see Appendix Table C.6).[20] We do, however, find some evidence of infant mortality reductions among the poorest ineligible households who at baseline possessed a non-thatched roof but walls and floors made of mud (Appendix Table C.7), which is consistent with the finding that a non-negligible share of the impacts on eligible households was driven by spillovers.[21]

The mortality effects estimated during the UCT period are broad-based across various pre-specified child and mother characteristics. As Appendix Figure C.2 reports, the impacts on mortality are nearly identical by child gender and similar by birth order. We find some suggestive evidence of heterogeneity by maternal age, with coefficient estimates about 75% larger in magnitude for older (aged 25 and over) than younger mothers, though both exhibit statistically and economically significant declines. As older mothers are more prone to birth complications triggered by strain (Yaman et al., 2025), this suggestive pattern is consistent with an important role being played by maternal health, which relates to the cause of death findings and evidence on mechanisms presented below.

Effects by Timing of Transfer Receipt

Precise date-of-birth data from the census, paired with administrative records of transfer disbursal from the GiveDirectly program, enable us to study differences in the effect of cash by timing of receipt. We find by far the largest treatment effects among children whose household was receiving cash in the month of their birth. Figure 3.2 illustrates treatment effects in high-saturation sublocations for the five birth timing groups noted above: (i) children whose household received cash 9 months to three years before their birth ("pre-pregnancy"), (ii) 9 months before birth to birth ("in-utero"), (iii) birth and up to 28 days after birth ("birth month"), (iv) 28 days after birth to one year after birth ("infant $\geq$ 1

[19] We do not estimate effects on under-five mortality for 2018-21 as children born after 2017 may not have reached their fifth birthday at the time of the census activity, as noted above.

[20] Results from Egger et al. (2022) and unpublished longer-term analyses show that living standards impacts of cash transfers persist in the medium-term, with consumption increases of 10-13% from 3 to 7 years post-treatment. However, recipients spend a large share of transfers in the first few months, and medium-term consumption gains are thus substantially smaller than in the immediate months after transfers (by a factor of roughly 5). If one were to assume a constant log-log relationship between spending and infant mortality, we would not be statistically powered to detect the mortality impacts of these far smaller changes in spending among cash recipients. The positive but smaller spillover consumption gains experienced by ineligible households would similarly imply smaller infant mortality gains that our design would, on average, be underpowered to detect. We did not pre-specify heterogeneity analyses within the ineligible population.

[21] Estimated spillovers for eligibles, as well as for the poor ineligibles, are not significant in the 2011-14 pre-UCT period, which serves as a useful balance check. In Appendix Table C.4, the spillover coefficient for eligibles in the pre-period estimated using Equation 3.1 is in fact positive in sign (+5.08, SE 4.79), whereas pre-period spillovers estimated using Equation 3.2 are similarly modest and not significant (-5.22, SE 5.01).

month"), and (v) one to three years after birth ("child $\geq$ 1 year"). We focus on reduced-form estimates since the sample sizes in some groups are modest; Appendix Figure C.3 reports results using the spatial IV, which yields similar results.

Infant mortality declines by 39.7 deaths/1,000 births among children receiving cash in their birth month ($p < .01$), suggesting that mortality fell to the low levels seen in industrialized nations. There is some evidence of mortality declines when cash is received in-utero (-13.1 deaths/1,000 births), although many children in this group were also exposed to cash during their birth month since the transfer window was 8 months. Strikingly, we estimate essentially no reduction in mortality in cases where a household received transfers before the pregnancy began. This result holds even if one considers the set of births where the last cash transfer was distributed 12 to 4 months before birth (before the last trimester of pregnancy), and we can reject the null hypothesis that the coefficient is equal to that of birth month receipt with 95% confidence (Appendix Figure C.4). Echoing the limited child mortality reductions after the UCT window, this result indicates that transfers are only effective when aligned with critical periods of pregnancy and a child's life, and suggests that households likely did not save a meaningful share of transfers, even when a pregnancy was already known. There are also limited effects when cash was distributed in a child's first year of life.[22] Moreover, and consistent with the absence of pre-treatment differences, there is no estimated effect of the UCT on infant (under-1) survival if a household received the transfer after the child's first year of life, which again serves as a useful specification check. Spatial IV results are similar, but with somewhat stronger evidence of effects for children exposed to cash in-utero or as infants (Appendix Figure C.3).

How is the apparent lack of savings which constrains effects of UCTs on infant mortality documented here reconciled with evidence that consumption is higher even 1.5 years on average after the transfers went out (Egger, Haushofer, Edward Miguel, Niehaus, et al., 2022b)? For one, the consumption effects dissipate over time, so households may lack resources for lumpy investments in health further removed from the transfers even if living standards are somewhat higher. Second, longer-run consumption gains may reflect investments in productive assets that modestly raise consumption but do not ease liquidity constraints as dramatically as UCT receipt. Third, the households with longer lasting consumption gains may differ in various ways from those that make health investments to reduce child mortality. In fact, we provide evidence below that child mortality reductions are concentrated among households with the lowest average 1.5 year consumption gains, consistent with this view.

In all, this stark heterogeneity by transfer timing indicates that factors present at the time of pregnancy and delivery and in the neonatal period may interact with cash receipt (in a setting with a high MPC) to produce particularly large declines in child mortality, and we return to these issues below.

[22] In Appendix Figure C.4, we additionally consider cases where the first transfer was received 4-12 months after birth and again find little evidence of a mortality reduction and can reject equality with the "birth year" coefficients. This result stands in contrast to the findings of McIntosh and Zeitlin, 2024, which estimated a 70% reduction in mortality among children already born in households that received UCTs in Rwanda (although that result does not survive multiple hypothesis corrections and is based on a very small sample).

Effects by Cause of Death: Verbal Autopsies

KGES field staff conducted verbal autopsies (VAs) using the World Health Organization's 2022 questionnaire to ascertain the likely cause of death for each of the 4,720 recorded under-five deaths in the household birth census. The field team was able to conduct VAs for 91% of child deaths across the period of study, and of these 82% were collected from a family member present at the time of the death, and thus is likely to be particularly knowledgeable about the circumstances. Following VA collection, a likely cause of death was assigned using the Institute for Health Metrics and Evaluation's SmartVA algorithm, which utilizes the Tariff 2.0 method for machine-learning classification of VAs and which was designed and validated with the Population Health Metrics Research Consortium Gold Standard VA database (Institute for Health Metrics and Evaluation, 2025).[23]

We focus on five pre-specified cause of death (COD) groups. The broad COD distribution seen in this study's VA data is similar to other studies in western Kenya (Amek, Odhiambo, et al., 2014). The leading COD group in control villages is maternal and neonatal causes, which encompasses individual causes such as death from preterm delivery, birth asphyxia, and congenital malformation. Maternal and neonatal causes comprise 37% of deaths with non-missing causes in low-saturation control villages. Pre-term delivery and stillbirths (if included) are the leading maternal and neonatal causes, accounting for 60% of this category. The second-largest COD group, encompassing 36% of deaths with non-missing causes, is communicable and nutritional diseases such as malaria and malnutrition. Other COD groups include respiratory diseases such as pneumonia (13%), non-communicable diseases (11%), and injuries (2%). A sixth category encompasses completed VAs for which the algorithm was unable to determine a likely cause due to missing or inconsistent answers (18% of completed VAs).

Figure 3.3 illustrates treatment effects on infant mortality by cause of death estimated using Equation (3.2) (here grouping together communicable/nutritional causes and respiratory causes, which in both cases are mainly due to infectious disease). The largest reduction in mortality by far is in deaths from maternal and neonatal causes: we estimate a drop of 11.4 deaths/1,000 births in this category, a 75% decline relative to the control mean.[24] Across other CODs, coefficients are almost always negative (the exception is non-communicable diseases, for which the point estimate is near zero), but mortality reductions across all other CODs combined amount to just half the decline seen within maternal and neonatal causes alone. We do not find evidence that VAs were differentially likely to be undetermined or absent in treatment villages, conditional on a death (Appendix Table C.9). We do find a drop in the overall death rates for undetermined or absent causes in treatment households, accounting for the remainder of the total mortality reduction.

[23]The likely cause of death is standardized, as defined by the International Classification of Diseases, tenth edition (ICD-10).

[24]This result is consistent with the finding that neonatal mortality, i.e., in the first 30 days of life, declines by 14.6 deaths/1,000 births in the census data, a 63% drop relative to the control mean (Appendix Table C.3). We find similar results for both the VA and neonatal mortality analyses when utilizing Equation (3.1).

These patterns confirm the central role of the birth event and the circumstances around it, and the importance of alignment of cash with birth timing. They also highlight the potential role of maternal health, a key correlate of neonatal deaths discussed further below.

Heterogeneity by Socioeconomic Status and Economic Treatment Effects

The hypothesized concave relationship between socioeconomic status and health would imply that socioeconomically deprived households would see the greatest reductions in infant and child mortality due to cash transfers. To test this, we make use of two sources of data: first, baseline values of total household assets from the household survey, and second, classifications of surveyed households as "most deprived" in terms of baseline consumption (which was not collected in the baseline KGES survey) from Haushofer et al., 2025.[25] More specifically, Haushofer et al., 2025 used detailed baseline household survey data from KGES to predict levels of per capita consumption using machine learning (ML) methods (generalized random forests). The bottom half of households were classified as "most deprived".

A related question is whether households with the largest positive treatment effects on medium-term economic outcomes (assets and consumption) experience differential impacts on child mortality. The hypothesis is that households in rural Kenya may face a trade-off between investing the cash transfers in child and maternal health versus investing in medium-term income generating activities or savings. Haushofer et al., 2025 classifies households by predicted treatment effects on assets and consumptions at endline (1.5 years on average after transfers) into those 'most vs. least impacted', again via ML methods.[26] They show that there is a strong correlation between being 'most deprived' and 'least impacted', i.e., the poorest households tend to gain least in terms of future consumption and assets.

As ML classification by deprivation and impact rely on data from the household survey, we restrict attention to eligible households in the household survey sample. We first benchmark the infant and child mortality estimates in this sample for reference by re-estimating effects restricting attention to individuals surveyed at baseline. Given the smaller sample size, these results are inherently less powered statistically. Despite this, the infant mortality estimates remain at least marginally statistically significant in the survey sample and somewhat larger in magnitude (reduced form effect of -36.7, p-value < 0.05) as compared to the overall census estimate (-25.3), see Appendix Table C.18 (column 1). The proportional reduction on the control mean is also almost identical to the census data, with the larger treatment effect being matched by a higher control, low-saturation infant mortality rate in the household survey subsample (65.8 per 1,000 births).

Figure 3.4 then presents heterogeneous treatment effects along the dimensions noted above: Panel A examines heterogeneity by deprivation, namely, the baseline value of assets

[25] This analysis was not pre-specified. We focus on assets and consumption since Haushofer et al., 2025 find the most evidence of predictable heterogeneity along these dimensions

[26] As the methodology generates multiple predictions for each household and outcome, we classify households as more deprived or impacted if the share of model runs that classify them as such exceeds the median.

and predicted per-capita consumption (absent treatment) from Haushofer et al., 2025.[27] In each case, we split the variable at the median, and report estimated coefficients for above-median (richer) households in red, and below-median (poorer) households in blue. The data present a consistent pattern of larger reductions for poorer households: by both measures, the point estimate for poorer households is larger in magnitude than that of the richer households and the differences are statistically significant at the 5% level. Strikingly, nearly the entire reduction in infant mortality appears concentrated among the poorer households.[28] This pattern also persists when controlling for the other dimension of heterogeneity. Estimates show evidence of substantial heterogeneity by deprivation even controlling for the corresponding impact classifications interacted with treatment assignment, and vice-versa. However, because the deprivation and impact measures are correlated, i.e., the most-deprived are more likely to be less-impacted (see Haushofer et al., 2025), standard errors increase somewhat when these controls are added (see Appendix Figure C.8).

These results provide some further support for the view that there is a concave relationship between socioeconomic status and health, and especially that improvements from very low living standards (as in rural Kenya) can be associated with pronounced health gains. This finding also provides another potential rationale for the lack of infant and child mortality effects among ineligible households, despite their documented gains from economic spillovers due to the cash transfer program: as the ineligible households have on average roughly twice the value of assets as the eligible households (and far lower baseline infant mortality), dramatic improvements in infant mortality may be more challenging to generate.

Figure 3.4 Panel B, shifts the focus to heterogeneity based on predicted impacts on endline 1 assets and consumption. This analysis indicates that infant mortality effects are largest in the groups with the smallest estimated economic impacts at endline 1. In fact, point estimates are essentially zero for the "most impacted" groups, and differences between the most versus least impacted are significant at the 10% and 5% levels for assets and consumptions, respectively. This result has two implications. First, this finding supports concerns that welfare measures based on consumption and assets typically omit non-monetary goods such as leisure or health that are challenging to measure (A. Deaton, 2016; Jones and Klenow, 2016), which may cause targeting policies to misallocate transfers towards households whose gains are easier to measure. Second, the fact that the households experiencing the largest reductions in child mortality had smaller medium-term economic gains suggests that they may have faced a stark trade-off between savings and investment in future economic gains versus investment into maternal and child health (possibly also through increases in leisure consumption by mothers shortly before and after birth, which we document below).

[27] We focus on reduced-form survey sample estimates using Equation (3.1) here as estimates using Equation (3.2) are somewhat less precise with the smaller survey sample when carrying out heterogeneity analyses.

[28] In Appendix Figure C.7, we report similar tests using predicted endline assets, which yields similar results, and by observed baseline and predicted endline income. Haushofer et al., 2025 find less evidence of predictable heterogeneity with respect to household income, which is more challenging to measure and fluctuates substantially in this setting. Consistent with heterogeneity being driven by true difference in household wealth, we find little evidence of heterogeneity along these income dimensions.

Taking the previous subsections together – on transfer timing, cause of death, and socioeconomic heterogeneity – suggests that returns to targeting cash could be particularly high if a policymaker were to target pregnant women from the poorest households. But in practice, pregnancy is a verifiable condition at relatively low cost, whereas it is likely far more expensive to identify more impoverished households in a rural East African setting like ours where subsistence agriculture and informal employment are widespread. Recall that all of the eligible household possessed easy-to-observe grass thatched roofing but the baseline value of household assets and income and the various living standards predictions require more time-consuming household surveys. This means it may be cheaper in practice to target on pregnancy status than on relative household poverty.

3.5 Comparison with Non-Experimental Variation

The birth census data enable us to benchmark the experimental cash transfer treatment effect estimates against several dimensions of non-experimental variation in economic circumstances. We provide a summary of these analyses below, with the details of each documented in Appendix C.2. In short, both the experimental and non-experimental estimates indicate that child survival is very sensitive to economic conditions in rural Kenya.

First, we examine the cross-sectional difference between transfer-eligible and ineligible households in control villages. Across 2015-17, infant mortality is 36% higher among control eligible births than among control ineligibles, a gap robust to controlling for basic birth demographic characteristics. Even within transfer-eligible households, substantial cross-sectional differences exist based on baseline household wealth: infant mortality is more than twice as high among households with below median baseline assets than among households above the median in control villages (Appendix Table C.11).

Second, we study the sensitivity of infant mortality to inter-temporal changes in economic conditions by comparing death rates in the pre-harvest "lean season," which is defined by Burke, Bergquist, and Edward Miguel (2019) as encompassing April through August, to the relatively prosperous harvest season. We find that across the pre-COVID period of 2011-19, rates of mortality for infants born in August (the peak of the lean season) are 21.9 deaths per thousand births higher than for infants born in the very next month (when the harvest has largely arrived), a difference similar in magnitude to the cash transfer treatment effect we estimate and representing a doubling of infant mortality for births a single month apart.

Third, we investigate how mortality responds to two major economic shocks which affected Kenya during the period the birth census spans: a major drought in 2017 and the COVID-19 pandemic in 2020-21. Infant mortality in the census sharply rises across the board during the 2017 drought as well as in 2020-21 (Figure 3.1, Panel A), and a regression discontinuity analysis indicates that the infant mortality rate doubled the week after Kenya imposed a strict COVID-19 lockdown on March 27, 2020 (Appendix Figure C.5, Panel A.).

Lastly, we move beyond the birth census and present the cross-country relationship between per capita GDP and infant mortality as a point of comparison (Appendix Figure C.5, Panel B). In 2014, each log point increase in per capita GDP was associated with a 0.79 log point reduction in infant mortality, a relationship which indicates that augmenting the study region's GDP by the same proportion as the UCTs did for treated households would predict a 36% decline in mortality. By comparison, the experimental estimates find a remarkably similar 44 to 48% decline.

In sum, across multiple sources of variation, infant and child mortality appears highly sensitive to economic conditions in this context, and non-experimental approaches recover similar magnitudes as the large experimental effects estimated in this study.

3.6 Mechanisms and Behavioral Change

The main finding of this study is that the disbursal of cash transfers leads to large reductions in infant and child mortality, and that this is concentrated among neonatal deaths. In this section, we turn to exploring potential drivers of these mortality declines. We do so by taking advantage of the multiple waves of detailed KGES household surveys conducted in the study region across over a decade, ranging from baseline data in 2014-2015 to the third endline survey round (2023-2025).

We begin by examining the selection of mothers into giving birth, and find that this cannot explain the main result. Second, we examine changes in healthcare utilization. We document large increases in hospital delivery and somewhat larger declines in child mortality for households near hospitals. Yet it is noteworthy that mortality reductions remain large in places far from hospitals where healthcare utilization does not increase. Healthcare utilization may thus amplify the impact of UCTs on child mortality but cannot fully explain them. We then turn to channels that are directly linked to maternal and child health: maternal labor supply, maternal self-reported health, and both maternal and child nutrition. We find large improvements in each measure consistent with these improvements playing a role in driving the mortality reductions.

Fertility Patterns

Prior research examining the effects of cash transfers has found that selection into fertility can play an important role in child health (e.g., Baird, McIntosh, and Özler, 2019). While we do find evidence that the cash transfers resulted in a modest transitory increase in fertility, features of the experimental design and analyses of the composition of women giving birth are inconsistent with fertility driving the reductions in child mortality observed in this paper.

First, we turn to the fertility effects of cash. We observe modest increases in births in the 2015-17 cash transfer period. Appendix Table C.10 reports estimates (from Equation (3.2)) indicating that among recipient households in high-saturation sublocations the share of women giving birth rose by 10% and births per 100 women grew 13% relative to the low-saturation control mean. Fertility patterns are indistinguishable by treatment status in the 2011-14 pre-period and again in the 2018-23 post-UCT period.[29] The impacts in the UCT period are transient: fertility remains unchanged when we look at the entire 2015-23 post-transfer period, and completed lifetime fertility among women who had reached age 45 in 2023 is similarly unmoved. Cash thus appears to have shifted the timing of births to a degree, but not increased the overall number of children born in the medium to long run.[30]

The timing of the transient fertility gains, however, does not align with the timing of the infant mortality reductions. As reported in Panel A of Figure 3.5, the rise in fertility is a gradual one. Consistent with the lack of selection bias, the treatment coefficient in the first nine months after transfers begin is not statistically significant and is in line with pre-period coefficients. It is only 27 to 45 months after transfers start that fertility exhibits a significant increase. By contrast, the fall in infant mortality occurs immediately after transfers commence, and in fact the largest estimated effect is observed in the first nine months following the start of transfers (Figure 3.5, Panel B). These births represent a useful test for selection effects: cash transfers were distributed over an eight month period, with households learning their treatment status less than a month before transfers started in their village. Mothers giving birth within 9 months of *any* transfers starting in the village therefore did not know their treatment status when pregnancy began, ruling out any endogenous selection into pregnancy. There are also significant reductions in infant mortality during the latter months of the cash transfer treatment period in which selection into fertility is a more distinct possibility.[31]

Several additional analyses indicate that differential selection into fertility is not a first-order determinant of child survival in this setting. We analyze the characteristics of women giving birth in 2015-17 in the representative survey sample. First, households with a birth during 2015-17 do not significantly differ by treatment status across six baseline socio-demographic characteristics: education, age, marital status, income, assets, and household size (Appendix Table C.15). Next, we use these six baseline household characteristics to predict the probability of giving birth over 2015-17 for all adult females in the census, and then examine whether treatment status is associated with the predicted probability of birth among women who gave birth during that key period. Birth probabilities are predicted using a random forest model trained with five-fold cross-validation. We find in Column 4 of Appendix Table C.16 that the predicted probability of birth among actual mothers does not significantly differ by treatment status in the transfer disbursal period (p-value = 0.47).

[29] Similar results are obtained using the reduced-form Equation (3.1).

[30] The fertility effects we estimate are not adjusted for age. Given the finding that fertility was shifted somewhat forward on average, controlling for age could introduce post-treatment bias, attenuating the estimated impact on fertility. Furthermore, the EL3 census did not collect detailed age data for all women who did not give birth, and hence we are unable to create fully age-adjusted fertility rates. We did collect full age data for all women who gave birth and therefore control for maternal age in the main child mortality regressions in an attempt to partly account for selection.

[31] Infant mortality remains lower by 10 to 17 per 1000 births — about a 30% reduction relative to the control mean — in the cash transfer period more than 9 after UCTs start, while fertility effects peak at 13%. Such a modest increase in fertility seems quantitatively unlikely to explain these large infant mortality reductions.

We further find (in Column 5) that treatment status is not associated with predicted birth probability among actual mothers in the post-transfer period of 2018-21 (p-value = 0.78). Predicting birth probability using LASSO from a set of 243 baseline household- and village-level characteristics, as we do in Appendix Table C.17, yields similar results indicating an absence of treatment impacts on the characteristics of women who gave birth.

Finally, we run Equation (3.2) for the representative baseline survey sample only, augmented with additional baseline controls to test whether the inclusion of household baseline characteristics affects the main results (Appendix Table C.18). The first column only includes pre-specified controls (e.g., year of birth fixed effects, birth gender, and mother age group), and there are sharp infant mortality declines due to cash transfers in this subsample, with an estimated drop larger than that found in the full census data (as noted above), at 55%. In Column 2, we then include controls for the six baseline socio-demographic characteristics previously considered, and the infant mortality results are again largely unaffected. In Column 3, we augment Equation (3.2) to include controls selected by post-double selection (PDS) LASSO (Belloni, Chernozhukov, and Hansen, 2014b), with a total of 243 baseline household- and village-level characteristics available for inclusion, and the results are also unaffected, with estimated drops if anything slightly larger (57%). Panel B of Appendix Table C.18 shows that these results are all robust to estimation using equation (3.1) rather than the spatial IV design.

Overall, selective fertility thus does not appear able to explain the large drops in infant and child mortality documented in this study.

Healthcare Access

We next examine whether the transfers allowed households to increase healthcare utilization. The endline 3 long-form household survey asked a representative sample of households detailed questions about their antenatal, delivery, and postnatal healthcare utilization during pregnancies which took place between 2015-17. Figure 3.6 reports estimated effects of the transfers on five pre-specified, WHO-recommended metrics of healthcare utilization (as well as on caesarean sections, C-sections, which were not pre-specified but are another potentially important channel), obtained from Equation (3.2).

The results indicate that mothers in eligible households were 20 percentage points more likely to give birth in a hospital if they received the UCT in a high-saturation sublocation, a large and statistically significant 45% increase (over the control mean of 44%).[32] These findings are consistent with past research which has argued that cost is a substantial barrier to institutional delivery in Kenya (Njuguna, Kamau, and Muruka, 2017). Hospital deliveries are particularly expensive: an assessment conducted shortly before the start of this UCT program indicates that the cost of delivering in a hospital stood at well over double that

[32] There is a small imbalance across treatment arms in the rate of hospital delivery in the pre-period (not shown). It is possible that some of this is driven by recall errors, for instance, if people mistakenly report the same delivery facility used during the cash transfer period for their earlier births. Recall that infant mortality rates are balanced in the pre-period (Appendix Figure C.1).

of non-hospital facilities (Institute for Health Metrics and Evaluation, 2014). For example, they find that delivering in a hospital cost \$137 (in 2011 nominal USD) on average, whereas delivering in a public health center cost \$56 and at a public dispensary just \$17.[33]

We cannot rule out that there were some negative congestion effects due to greater use of local hospitals for deliveries, as some recent research suggests could be relevant in other LMIC settings (Andrew and Vera-Hernández, 2024). However, we tabulate only 4.3 births per week at the average hospital in our study area during the 2015-17 period, suggesting limited scope for congestion. Moreover, the results show that any such congestion effects are smaller than the gains experienced by cash transfer recipients. Specifically, we do not document any adverse net impacts of the program among ineligible households who did not receive a transfer (Appendix Table C.6). In fact, Appendix Table C.7 documents that the poorest ineligible households — who experienced more modest but still sizable reductions in infant mortality, as noted above — were also able to deliver in hospitals at a significantly higher rate during the cash transfer period.

Does the expansion in institutional delivery alone explain the large reduction in child mortality? To further examine the potential role of healthcare access, we assemble detailed data on travel times to health facilities across the study region. The coordinates of every registered health facility nationwide were obtained from the Kenya Master Health Facility Registry (KMHFR; Republic of Kenya Ministry of Health (2025)). We then surveyed the facilities to collect information about the care that they offer, including whether the facility is staffed by a physician. We augment this data using surveys of health facilities that household respondents had reported visiting but were not present in the KMHFR data. Field officers equipped with GPS speedometers logged the travel speeds of over 1,100 trips throughout the study area to construct a network of actual travel times on roads. We then complement this with estimates of walk times off of roads using a procedure previously utilized in the study region (Ouko et al., 2019) to compute total travel time to each facility.

Table 3.2 column (1) shows that reported increases in hospital delivery are concentrated among households with below-median travel time (mean travel time 23 minutes) to such a facility, with no significant effect in areas far from hospitals (mean travel time 55 minutes). If hospital delivery were the main driver of lower mortality, one would therefore expect mortality reductions to likewise be concentrated in areas near hospitals. But Table 3.2 column (3) shows that there is a large and statistically significant reduction of 20 infant deaths per 1,000 births even in areas that are farther from hospitals. Panel B further selects regression controls using double partial-out LASSO to account for omitted variable bias (confounding) from place-based factors that might be correlated with the placement of health facilities — including a vector of measures including population density, distance to

[33] In terms of other health care utilization outcomes, the table shows that cash transfers do not significantly affect the proportion of pregnancies where the WHO recommended number of at least four antenatal visits occurred. Point estimates also suggest a decline of roughly half in C-section delivery rates, although differences are not statistically significant. Cesarean delivery is positively correlated with mortality in control areas, suggesting that the observed decline in rates may reflect a reduction in emergency surgeries ("crash C-sections") among cash recipients.

towns and roads, baseline village wealth, malaria suitability, and rainfall, among others (and interactions of these covariates with the treatment variables). There is stronger evidence that access to care may increase effect sizes in this specification, although differences are not significant. We repeat this analysis using distance to a physician-staffed facility (which Okeke (2023) suggests is particularly important to birth outcomes). Effects remain large and significant far from such facilities, although the estimated interaction with controls in Panel B again suggests that effects are stronger when households can more easily access care ($p < 0.1$).

The evidence therefore suggests that healthcare access can be complementary to UCTs and lead to larger child survival gains, but that the effect of transfers remains large and significant even when such resources are not locally available or utilized. Other mechanisms thus have to explain a meaningful share of the reductions in child mortality, and we turn to some of these next.

Maternal Labor Supply

Above, we showed that infant mortality reductions are largest among households that experienced the smallest monetary gains over the 1.5 years after cash transfer receipt. One explanation consistent with this result is that households may have traded off investments into child health with those into future income-generating activities. We test this prediction by examining changes in maternal rest before and after birth. Parental time is believed to be a key input in the production of child health (Ruhm, 2000; G. Miller and Urdinola, 2010; Rossin, 2011; Nandi et al., 2016; Bartel et al., 2023). Many determinants of child health, such as traveling to primary care visits and rest during pregnancy, are inexpensive monetarily but time-consuming, creating a high opportunity cost. Kenya is in the top ten countries worldwide for female labor force participation (at 72%), and of all global regions, Sub-Saharan Africa features the highest rates of women working (International Labour Organization, 2025). While in general high female labor supply can have a number of positive effects on the well-being of women and children (Heath and Jayachandran, 2018), performing strenuous physical tasks for extended periods of time during pregnancy and the initial months postpartum, as is common in Kenya, may have deleterious consequences for infant health (Izugbara and Ngilangwa, 2010; Riang'a, Nangulu, and Broerse, 2018; Scorgie et al., 2023). As background on the study setting, women in control villages engage in nearly 40 hours of productive work on average per week in the last trimester and first three months after birth.

Egger, Haushofer, Edward Miguel, Niehaus, et al., 2022b document little effect of the transfers on average household labor supply, with positive point estimates statistically indistinguishable from zero. Families with a woman in late-stage pregnancy or with a newborn, however, may exhibit different patterns. For women in these cases, the cash arrives at a time when the marginal utility of rest and time to invest in child health may be particularly high. Figure 3.7 reports labor supply impacts of cash separately by gender, based on Equation (3.2) and including treatment terms interacted with indicators for a pregnancy or newborn present at the time they were surveyed at endline 1.[34] In separate regressions, we include indicators for three periods of interest: the first six months in-utero, the third trimester in-utero and first three months postpartum, and the next six months postpartum (four to nine months after birth).[35]

We find substantial heterogeneity in labor supply effects among women by whether a late-stage pregnancy or newborn is present in the household. In line with Egger, Haushofer, Edward Miguel, Niehaus, et al., 2022b, recipient households without a pregnancy or recent birth exhibit no change in hours of labor supplied: for both women and men, the point estimate on weekly hours is not statistically significant and close to zero. In the three months before and after a birth, however, cash transfers reduce female labor supply in high-saturation recipient households by 24.48 hours a week, a 53% decrease relative to the control group mean ($p < .05$). By contrast, among men we estimate effects that are near zero and not significant across all periods, suggesting that it is women in particular who are able to temporarily reduce labor supply when a pregnancy is present due to the transfers.[36]

These results provide evidence that the cash transfers may have enabled some women to reallocate time from potentially strenuous labor to rest or other activities more beneficial for fetal and child health. Reductions in labor supply during pregnancy as a channel lowering infant mortality is consistent with prior work on the importance of parental time in the production of child health, and accord with studies in other LMIC settings indicating an association between cash transfers and reductions in maternal labor supply (Novella et al., 2012; Amarante et al., 2016; Garganta, Gasparini, and Marchionni, 2017; Guldi et al., 2024). It is also consistent with evidence of extensive labor supply during women's final trimester of pregnancy likely resulting in increases in mortality that align closely with the causes of death we identify. Cai et al., 2020 report the results of a meta-analysis finding that lifting objects, physically demanding labor, and prolonged periods of standing significantly increase the odds of preterm delivery, low birth-weight, and pre-eclampsia. The health literature also documents that excessive work hours, such as those reported in our setting, are a particularly high risk factor for neonatal mortality. For instance, Kader et al., 2021 find that women who worked one 40 hour week during their third trimester of pregnancy doubled their risk of a preterm delivery, with additional risk from frequently working long hours.

In all, the large decline in work hours documented in this study, generally in physically demanding professions, from approximately 40 hours to around 20 hours per week, may thus plausibly translate into meaningful reductions in child mortality.

[34]This analysis was not pre-specified.

[35]See Appendix Table C.12 for the estimates from Figure 3.7 in table format. Estimates encompass only households selected for long-form surveys which inquired about labor supply; as such, the sample size is smaller than the full birth census. Six-month bins are thus used to increase statistical power.

[36]Further heterogeneity analysis based on the gender of the cash recipient within the household could be valuable. However, we observe only the name of the individual who registered the cellphone that transfers were sent to. Cellphones are often shared by households and GiveDirectly did not have a policy to target transfers by gender, so we cannot reliably determine if the actual cash recipient was female.

Maternal Health and Nutrition

Lastly, we examine the effects of transfers on the health of mothers and the nutrition of children. Measures of adult health from the first endline survey conducted in 2016-17 are consistent with recent and expectant mothers benefiting most from the cash transfers. Adult health was measured using a pre-specified standardized average of three components: an index of recent symptoms for health conditions (e.g., fever, persistent cough, blood in stool, stomach pain), a binary variable capturing whether the respondent reported experiencing a major health problem that affected their work or life since the baseline survey, and the respondent's self-reported level of current general health rated on a scale of one to five.

Across the general population of adult recipients, estimates of treatment effects are close to zero and not statistically significant, as reported in Egger, Haushofer, Edward Miguel, Niehaus, et al. (2022b). As with labor supply, however, treatment effects are different for new mothers. Among female survey respondents in a household with a child aged one or younger, column 1 of Appendix Table C.13 reports that UCTs increased the health index by 0.28 standard deviation units (p-value < 0.05). In contrast, we find no changes in physical health for eligible men with infants in the household nor eligible female respondents without infants.

An examination of the components of the health index suggests that the increase in the index is driven by fewer major adverse health episodes as opposed to current perceptions of health. Appendix Table C.14 shows that transfers reduced the share of recipient women with infants having experienced major health problems that seriously affected their life or work since baseline by 9.1 percentage points, a 66% decrease over the control mean of 13.8 percentage points. In contrast, self-reported current health is virtually unchanged, though the point estimate is positive. The third component of the index, which is itself a standardized index of health symptoms in the last four weeks, exhibits a 0.22 standard deviation unit decrease (p-value < 0.05). Last, column 5 of Appendix Table C.14 shows that conditional on having experienced a major health episode, transfers do not render it more likely that the health problem was resolved. Rather, transfers reduce the likelihood there was a major health problem in the first place.

Another important consideration is the role of maternal and child nutrition. The first endline survey conducted in 2016-17 indicates that food security for children significantly rose for recipient households: a pre-specified index of child food security, which encompasses household survey questions on whether children skipped meals, went to bed hungry, and went entire days without food over the past week, increased by 0.18 standard deviation units on average among recipient baseline households. This result, which is reported in Appendix Table C.13, is statistically significant at the one percent level. The final column of Appendix Table C.13 reports a marginally significant positive rise in the analogous adult food security index among recipient baseline households of approximately 0.10 standard deviation units.[37]

[37] Estimates are of a similar magnitude among recipient women with an infant present in the household, though we lack the statistical power to examine these subgroup effects with much precision.

These nutrition gains, as well as the reductions in maternal labor supply, do not necessarily depend on improved access to healthcare. These mechanisms are therefore consistent with the finding that infant and child mortality reductions remain large even absent observable increases in healthcare utilization (i.e., among households living far from a hospital). UCTs thus may yield meaningful benefits even in settings where such infrastructure is limited if they can directly improve the health and nutrition of mothers and children.

3.7 Cost-Effectiveness Implications

The prior discussion of potential mechanisms prompts two important questions. First, what do the estimated child mortality benefits imply about the welfare gains from UCTs versus other forms of assistance? Second, are cash transfers a cost-effective tool for reducing child mortality rates, and can the mechanism analyses reported above help guide efforts to target transfers to those most likely to benefit?

We estimate the number of child deaths averted among recipient households during the transfer disbursal period using the following "back-of-the-envelope" calculation:[38]

$$\begin{aligned}&(\text{Estimated average treatment effect on recipient child mortality})\\&\times(\text{Number of births among recipient treatment households during 2015-17})\,.\end{aligned} \tag{3.5}$$

The estimates reported in Table 3.1, Column 4 reflect the average total effect in high-saturation sublocations. To estimate lives saved, we therefore apply an estimate of the average treatment effect on child mortality pooled across both high and low-saturation sublocations (instead of focusing on high-saturation cases alone), which is -24.19 (SE 7.81). There were 3,533 births for eligible households in treatment villages across 2015-17, so we estimate that approximately (-24.19/1000) x (3,533) = 86 child deaths were averted due to the UCTs. This is a substantial reduction: based on the low-saturation control village mean of 57.4 deaths per thousand births, we would have expected approximately 203 child deaths to have occurred in treatment villages across this period.

The leading approach to estimate the welfare gains of mortality reductions is by recipients' value of a statistical life (VSL). Revealed preference estimates of VSL, consumers' demand for mortality risk reduction, tend to be low among populations with income levels similar to this study. In fact, most revealed preference studies we are aware of find values below $5,000 (Killeen, 2025; Kremer, Leino, et al., 2011; Berry, Fischer, and Guiteras, 2020). As documented in Killeen (2025), economic theory only supports applying these VSLs to balance the trade-off between consumption and aid to reduce mortality; low values do not imply that resources dedicated to improving health should not be allocated to poor households. This level of VSL would imply welfare gains from the mortality reductions below

[38] We focus on child mortality as opposed to infant mortality due to its relevance for policymakers and foundations. For example, the United Nations Sustainable Development Goals refer specifically to under-five mortality but not to under-one mortality (United Nations, 2015).

$500,000. However, others have argued for higher VSLs in low income settings. For instance, GiveWell, a non-profit charitable giving advisor applies "moral weights" which (based on our understanding) value averting an under-5 death at 116 times the benefit of doubling annual consumption, implying a value of $87,956 per life saved in this setting (and aggregate gains of USD 2023 PPP 7.6 million).

To account for the wide range of VSL estimates, we report the estimated benefits of a $1,000 investment in UCTs (including the multiplier gains documented in Egger, Haushofer, Edward Miguel, Niehaus, et al., 2022b), versus a leading health intervention, malaria medication, by VSL in Figure 3.8.[39] Four VSL estimates obtained from settings with similar income levels are included in the lower panel. We focus on the range of revealed preference estimates since they are based on choices with real stakes, but we additionally include a stated-preference estimate from Redfern et al. (2019), which estimates a VSL of over $55,000, because it informed GiveWell's moral weights and is used in important policy decisions.

In the top panel, we present the estimated welfare gains of UCTs across three different scenarios and contrast them to the gains from the malaria treatment intervention. The first "base" UCT case excludes any benefits from economic spillovers or child lives saved, and thus values the $1,000 transfer at exactly $1,000 (the horizontal green line). In this case, the malaria intervention generates larger welfare gains than cash transfers even at relatively modest levels of the VSL, as the green and red lines intersect at approximately $4,000.

The second case, denoted by the thin blue line, includes the benefits from both the economic spillovers and child lives saved documented in this study. The previously presented results suggest that such gains largely accrue in different households: those with the largest mortality reductions are those with the weakest evidence of economic gains. We thus interpret these figures as the average gain across households. Recall that Egger, Haushofer, Edward Miguel, Niehaus, et al., 2022b estimate a real transfer multiplier of approximately 2.5 in the study area, leading to an increase in welfare of approximately $2,500 at low VSLs.[40]

The welfare gains of a UCT program like the one that we study are mainly driven by consumption gains, and there is a proportionally small marginal gain in welfare when accounting for mortality reductions across most VSLs in the distribution. This is true because the UCTs were not targeted to pregnant women, so the cost per life saved is relatively high.

Third, we plot estimated welfare gains if the UCTs were targeted to pregnant women (retaining the assumption of a transfer multiplier of 2.5), in the thick blue line.[41] The welfare gains from mortality reductions are much larger in this scenario given that far more births are affected by the transfer: targeted cash transfers yield far higher estimated welfare gains

[39]We use GiveWell's estimate of the cost per life saved through the malaria intervention of $4,304. We selected this program because it was GiveWell's top listed charity at the time of writing in March 2025. This estimate accounts for the spillover benefits of malaria treatment.

[40]Egger, Haushofer, Edward Miguel, Niehaus, et al. (2022b) note that interpreting the transfer multiplier as welfare gains can be problematic if factors such as reduced leisure or savings drive consumption gains, however, they find no evidence of such responses, so we assume the transfer multiplier translates into welfare gains for this analysis.

[41]Here we abstract away from any potential fertility responses to a targeted UCT program.

than untargeted transfers for VSL values above about $20,000.

Across the three cases we consider, the welfare gains from cash transfers are larger than those from malaria medicine for low levels of the VSL corresponding to most of the existing revealed preference VSL estimates (below roughly $4,000). However, if one values lives saved by the far higher stated preference estimate of $55,000, then the malaria intervention generates larger welfare gains than any of the UCT estimates. The welfare gains from a UCT program targeted to pregnant women are greater than those generated by the malaria treatment program up to a VSL level of approximately $11,500.

We next add structure to the problem and estimate the posterior distribution of the VSL in this population using Bayesian hierarchical meta analysis. The plot reveals that the estimated welfare gains of the targeted UCT program dominate malaria medicine for about 75% of the distribution of VSLs, although the mean estimate of gains from malaria medicine are $300 higher. This holds because the Redfern et al. (2019) VSL estimate induces a rightward skew in the distribution. Both the untargeted UCT program and the targeted UCT program yield higher estimated welfare for the full 95% confidence interval of VSLs if only revealed preference VSL studies are examined. Thus in cases where UCTs produce the general equilibrium effects documented in Egger, Haushofer, Edward Miguel, Niehaus, et al. (2022b), we view UCTs as an attractive form of aid for a wide range of plausible VSL values, especially when they are targeted to pregnant women, even in comparison to highly cost effective health interventions. Details of the estimation of the VSL distribution and the welfare analysis are in Appendix C.3.[42]

A second question related to the welfare implications of the child mortality reductions is how cost-effective UCTs are compared to a range of other health interventions while focusing more narrowly on child survival (and excluding consumption gains). As documented in Killeen (2025) and noted above, the prior welfare analysis guides optimal policy if donors are deciding between various programs to benefit a fixed population, but if funds are specifically earmarked by donors for health, economic theory does not support the use of recipients' VSL. We therefore benchmark the cost per death averted to other health interventions in Sub-Saharan Africa considered cost-effective by the World Health Organization.

In total, the UCT program under study disbursed USD PPP 25.75 million, and given that we estimate the program averted approximately 86 child deaths, this implies a cost of USD PPP 299,418 per death averted. However, as discussed above, the UCTs were primarily intended to raise consumption and not targeted to pregnant women. We therefore consider targeting transfers to households with women in the third trimester of pregnancy. Disbursing UCTs to these households would cost a total of USD PPP 1.65 million (USD 700,000 nominal) in the study sample and time period, based on the household survey data.

Calculating the number of deaths averted under this scenario is challenging due to two opposing factors. On one hand, restricting transfers to a subset of households reduces the

[42] Appendix Figure C.9 reports the results of a decision theoretic model which yields similar results. Namely, broadly targeted UCTs minimize median regret, but malaria medicine narrowly minimizes Bayesian regret when Redfern et al. (2019) is included in posterior estimates of the VSL.

spillovers from other treated households. But targeting cash to particular subpopulations may result in larger treatment effects among those high-impact groups. We found earlier in Figure 3.2, for example, that mortality was virtually eliminated for children whose households were receiving the UCT in the month when they were born, in contrast to other timing cases in which there were more modest effects. To be conservative, we therefore utilize the average high-saturation treatment effect for all recipients (-25.63 deaths/1,000 births).

Targeting UCTs to women in the third trimester of pregnancy under these assumptions would cost about USD PPP 92,000 (or $39,000 in nominal dollars) per child death averted. We benchmark these calculations to 37 WHO-recommended maternal and child health interventions in East Africa as estimated by Stenberg et al. (2021). Across interventions and scenarios, the cost per death averted ranges from USD PPP 27 to USD PPP 222,952.[43] Hence, even without taking into account any of the other documented benefits of UCTs (such as gains in consumption), the transfers are squarely in the range of cost per death averted among these WHO-recommended interventions.

One could also imagine targeting UCTs based on baseline socio-economic status or predicted impact, as results show that mortality gains are concentrated among households with the lowest consumption levels and smallest consumption gains. For instance, the estimated cost per life saved among households with below median household consumption is about USD PPP 33,000 ($14,000 in nominal dollars). However, while such targeting efforts may be cost-effective, they depend on rich data that is not typically readily available.

3.8 Conclusion

A large-scale unconditional cash transfer program in rural Kenya led to a sharp drop of nearly one half in infant mortality. The largest mortality reductions were observed among households that received cash near a child's birth month, in poorer households, and among households with smaller economic gains from the transfers in the medium term. Concomitant with large mortality reductions, cash leads to substantial reductions in maternal labor supply, improvements in mothers' health, and nutritional gains for mothers and children. We find evidence that access to higher quality health facilities like hospitals may amplify these gains, but there remain sizable reductions in mortality even among households living far from health infrastructure. A rough calculation suggests that transfers targeted to pregnant women are similarly cost-effective to a number of child health interventions recommended by the WHO. These child mortality reductions represent benefits of UCTs beyond the direct and spillover household consumption gains already documented (Egger, Haushofer, Edward Miguel, Niehaus, et al., 2022b).

The large magnitude of the child survival gains underscores the fact that infant and child mortality appear sensitive to economic conditions in low-income contexts, such as the rural

[43] Stenberg et al., 2021 evaluates cost-effectiveness using three coverage level scenarios: 50%, 80%, and 95%, and report health impacts in terms of healthy life years (HLY) saved. We converted HLYs to deaths averted using WHO data on total and healthy life expectancy in Kenya (World Health Organization, 2025).

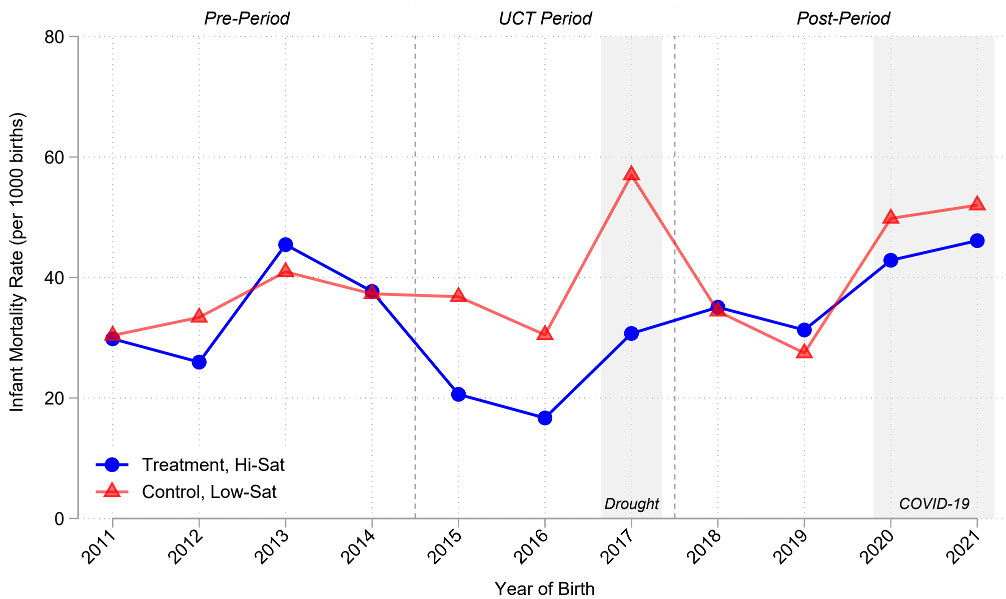
Kenyan study setting. The socioeconomic gradient in mortality is steep in rural Kenya: not only do large UCTs nearly halve infant mortality, but treatment effects are concentrated among the poorer households in the sample, and mortality rates vary substantially by household baseline wealth in the cross-section, as well as inter-temporally by the agricultural harvest season.

An important implication of these results is that cash can be effective at reducing infant and child mortality even absent health infrastructure because it improves the health of mothers and children through channels such as increased maternal rest and improved household nutrition. This suggests that such programs may yield meaningful child survival gains in other low-income settings that have more limited healthcare infrastructure than Kenya. The timing of transfers is also highly consequential: households receiving cash near the time a child is born see by far the largest resulting reductions in mortality. Outside of public health policy, this finding supports the view that allowing households to select the distribution time of cash transfers may increase their efficacy by aligning available liquidity with high-return investment opportunities (Kansikas, Mani, and Niehaus, 2025).

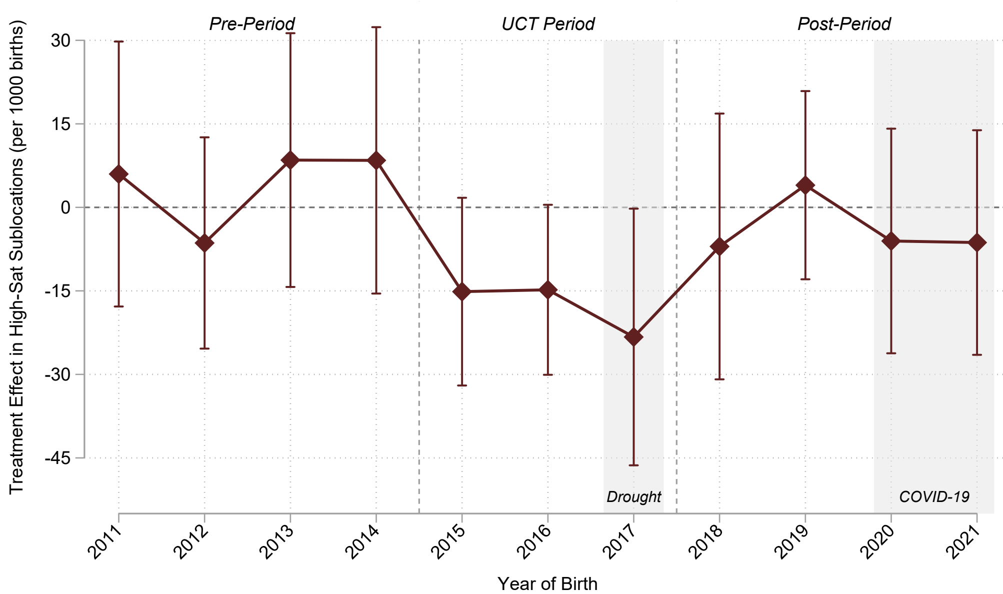
Finally, this paper suggests that typical measures of consumption, which focus on market goods that are relatively easily measured, may miss important welfare gains from the consumption of non-market amenities such as leisure and health. Economists have long noted that consumption of such amenities is frequently omitted from consumption surveys and GDP. We find that child mortality reductions are concentrated among the households with the smallest documented consumption gains following cash transfer receipt, suggesting that poor households may face a stark trade-off between monetary consumption versus non-monetary investments in health.

3.9 Tables and Figures

Figure 3.1: Unconditional Cash Transfers and Infant Mortality By Year



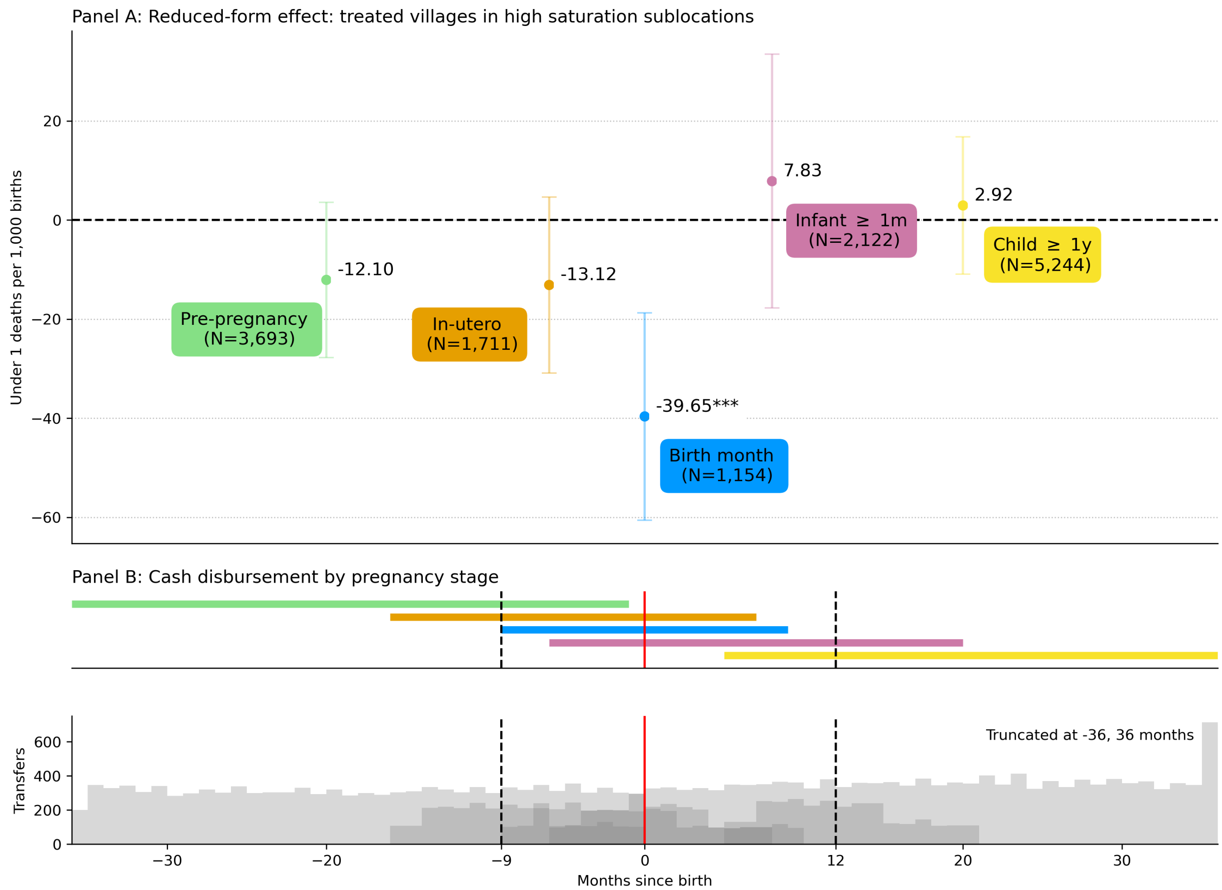
Panel A: Infant Mortality By Year



Panel B: Reduced-Form Impacts by Year

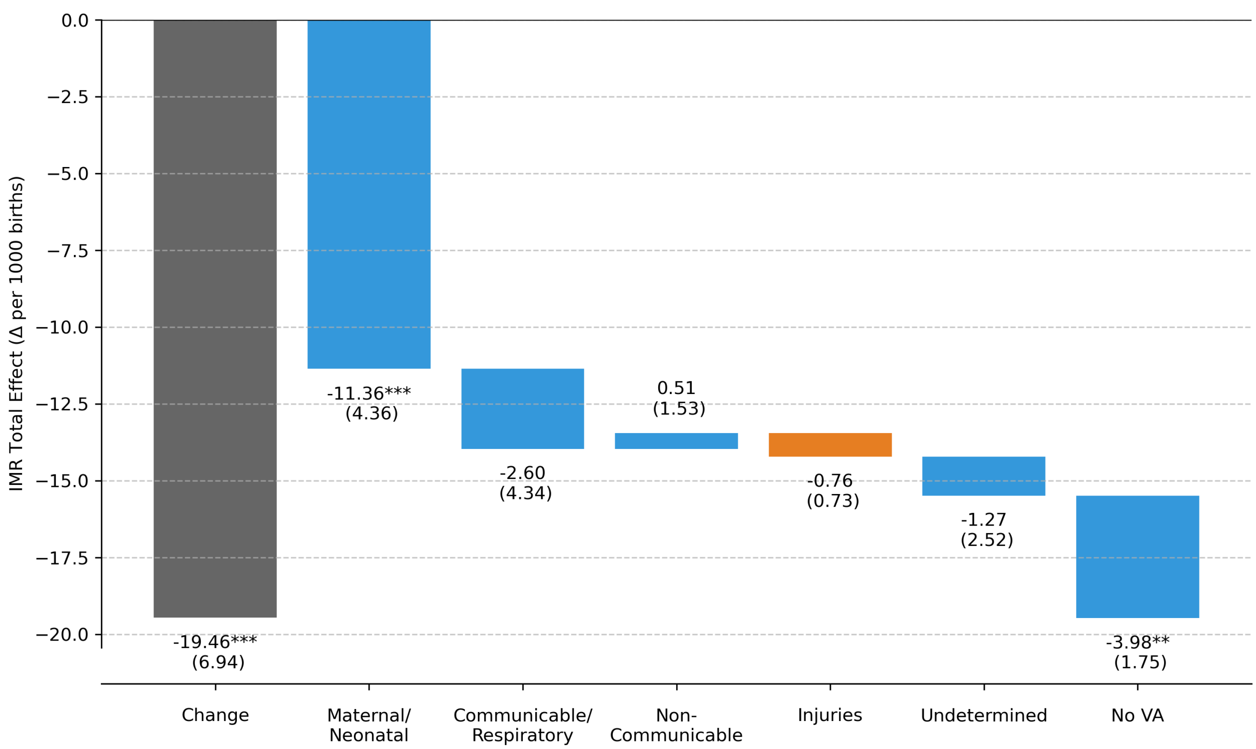
Notes: This figure is based on the main EL3 birth census sample encompassing births from 2011 to 2021. Panel A reports the mean infant mortality rate by year among eligible households for treatment villages in high-saturation areas and control villages in low-saturation areas. Panel B reports the year-by-year reduced-form estimates of infant mortality impacts among eligible households for treatment villages in high-saturation areas. Pre-period births refer to those occurring in the period 2011-14, whereas the unconditional cash transfer (UCT) period refers to 2015-17 and the post-UCT period denotes 2018-21. The COVID-19 pandemic spans 2020-21 and a drought affected Kenya in late 2016 and 2017. The whiskers on each yearly estimate denote the 95% confidence interval. Standard errors are clustered at the sublocation level.

Figure 3.2: Transfer Effects on Infant Mortality by Timing of Cash Disbursal



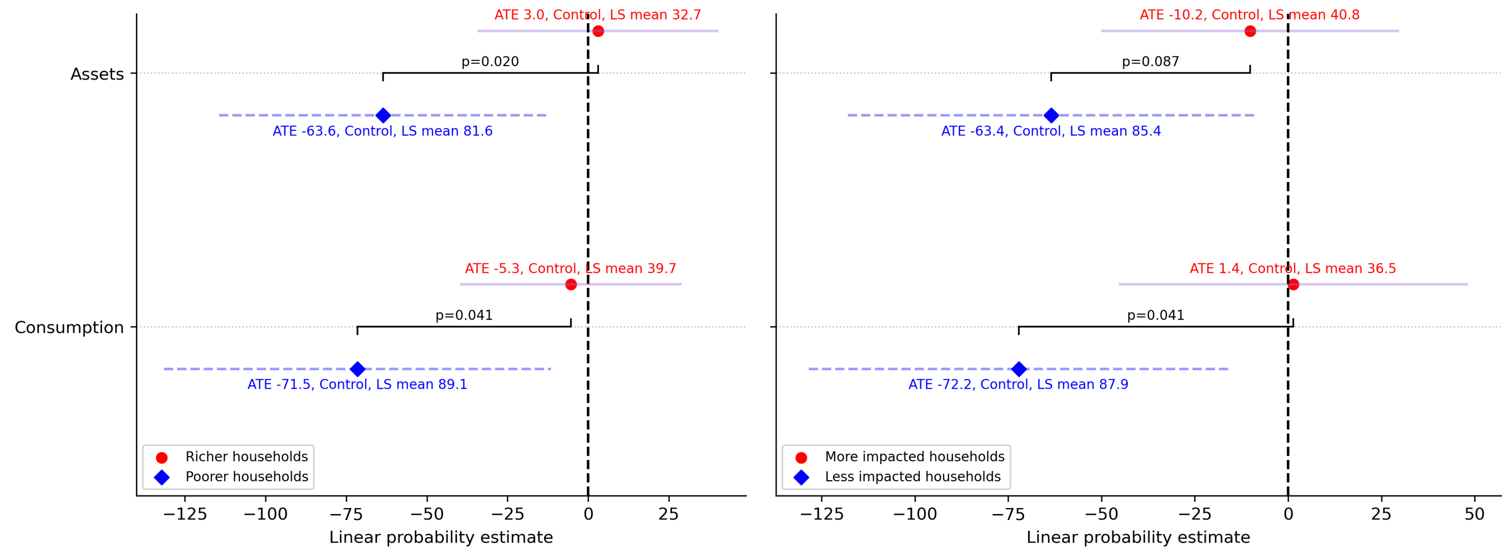
Notes: This figure is based on the main EL3 birth census sample. Panel A plots estimates of dynamics based on the time period of exposure to cash during the child's life. The transfer timing is defined relative to the "experimental start date", as this is well-defined for both treatment and control villages. "Pre-pregnancy" includes household exposure 3 years to 10 months before birth. In-utero is 9 to 1 month before birth. Birth month includes cash within the first month of life. Infant includes 1 month to 12 months. Child includes 1 to 3 years (and can be viewed as a placebo check on infant mortality). Estimates are constructed using equation (3.4), which estimates equation (3.1) after restricting the sample to those exposed to cash at a particular time relative to the birth month. Observations appear in multiple groups since cash transfers were distributed over 8 months, and we include all observations where the exposure period overlaps with this 8 month window. Panel B plots the range of experimental start dates, relative to birth month, included in each estimate and a histogram of transfers by month in each bin. The spatial IV version of this figure (which is estimated using Equation 3.2) is presented as Appendix Figure C.3. 95% confidence intervals are shown. * $p < .10$, ** $p < .05$, *** $p < .01$.

Figure 3.3: Transfer Effects on Infant Mortality by Likely Cause of Death, 2015-17



Notes: This figure is based on the main EL3 birth census sample encompassing births from 2015-17 and reports the estimated reduction in infant mortality (per 1000 births) based on cause of death determinations from verbal autopsies (VAs). Treatment effects are estimated using Equation (3.2). The control, low saturation mean rates per 1000 births for the categories are Maternal/Neonatal: 15.20, Communicable/Respiratory: 10.30, Non-communicable: 2.45, Injuries: 0.49, Undetermined: 7.36, No VA: 4.41, Overall: 40.21. The undetermined category encompasses completed VAs for which the SmartVA algorithm was unable to determine a likely cause due to missing or inconsistent answers. The no VA category includes cases for which no VA was collected. * $p < .10$, ** $p < .05$, *** $p < .01$.

Figure 3.4: Heterogeneous treatment effects on infant mortality by deprivation and impact (reduced-form)

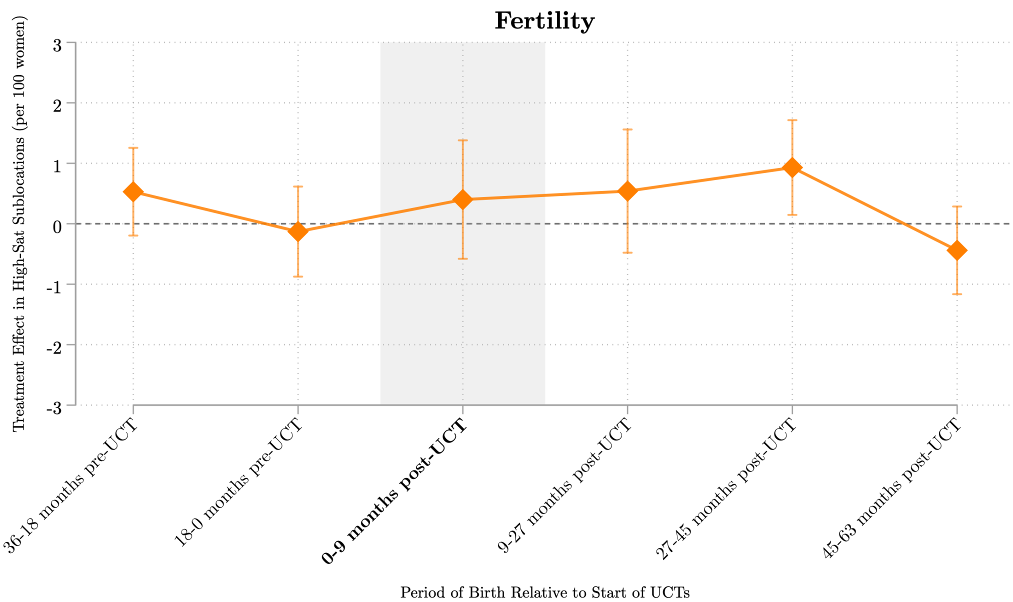


Panel A: Heterogeneity by baseline socioeconomic status

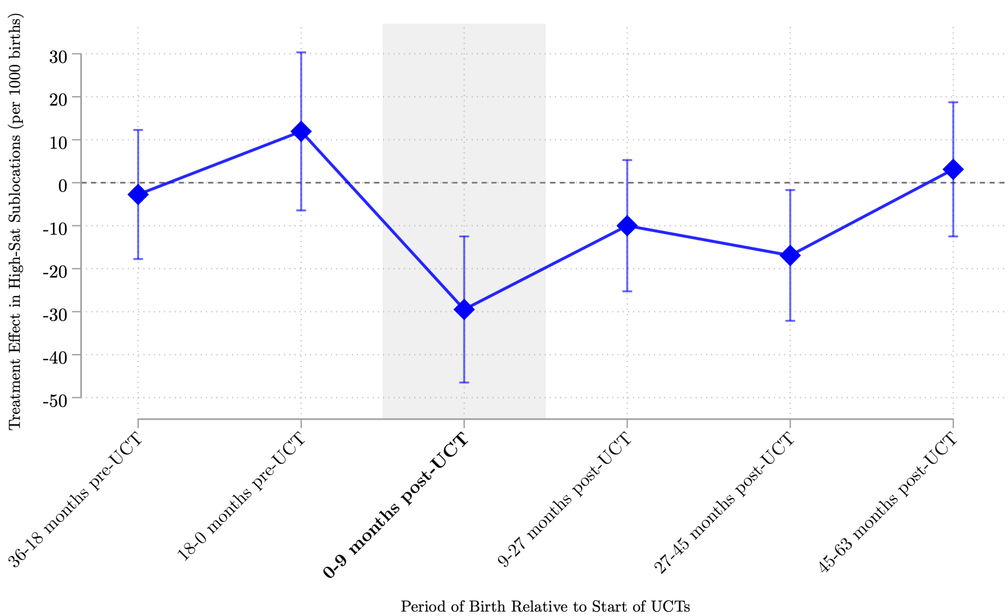
Panel B: Heterogeneity by predicted conditional average treatment effect (CATE)

Notes: Transfer-eligible households visited in the baseline survey and EL3 census are included. Panel A reports effects on infant mortality (per 1000 births) by below vs above-median wealth, estimated using Eq. (3.1). In Panel A, richer households by the measure of "assets" reported above median baseline household assets, while "consumption" uses the predicted endline per capita generalized random forest (GRF) prediction from Haushofer et al. (2025). Consumption is predicted since it was not measured at baseline. A household is defined as "richer" if the share of GRF runs where it was defined as "most deprived" per the Haushofer et al. (2025) was lower than average. Panel B reports effects on infant mortality (per 1000) births based on whether the predicted conditional average treatment effect (CATE) of per capita assets or consumption was lower or higher. Specifically, a household is defined as "more impacted" if the share of GRF splits in Haushofer et al. (2025) that classified it as "most impacted" exceeds the sample median, across the full sample of births among eligible HHs. Income is excluded from this figure since Haushofer et al. (2025) find less evidence of predictable heterogeneity on that dimension. 244 observations fall in above median BL assets and predicted asset CATE, 335 had below median assets but above median predicted CATE, 209 had above median BL assets but below median predicted CATE, and 473 were below median on both dimensions (correlation = -0.1). With respect to consumption, these counts are 438, 137, 179, and 507 respectively (correlation = -0.5). Stars denote the significance of the difference between richer/poorer or less/more impacted households. Figure C.7 reports effects by income and predicted (rather than observed) baseline assets. P-values on the probability of equality of the groups are plotted in black.

Figure 3.5: Transfer Effects on Fertility and Infant Mortality by Timing of Birth



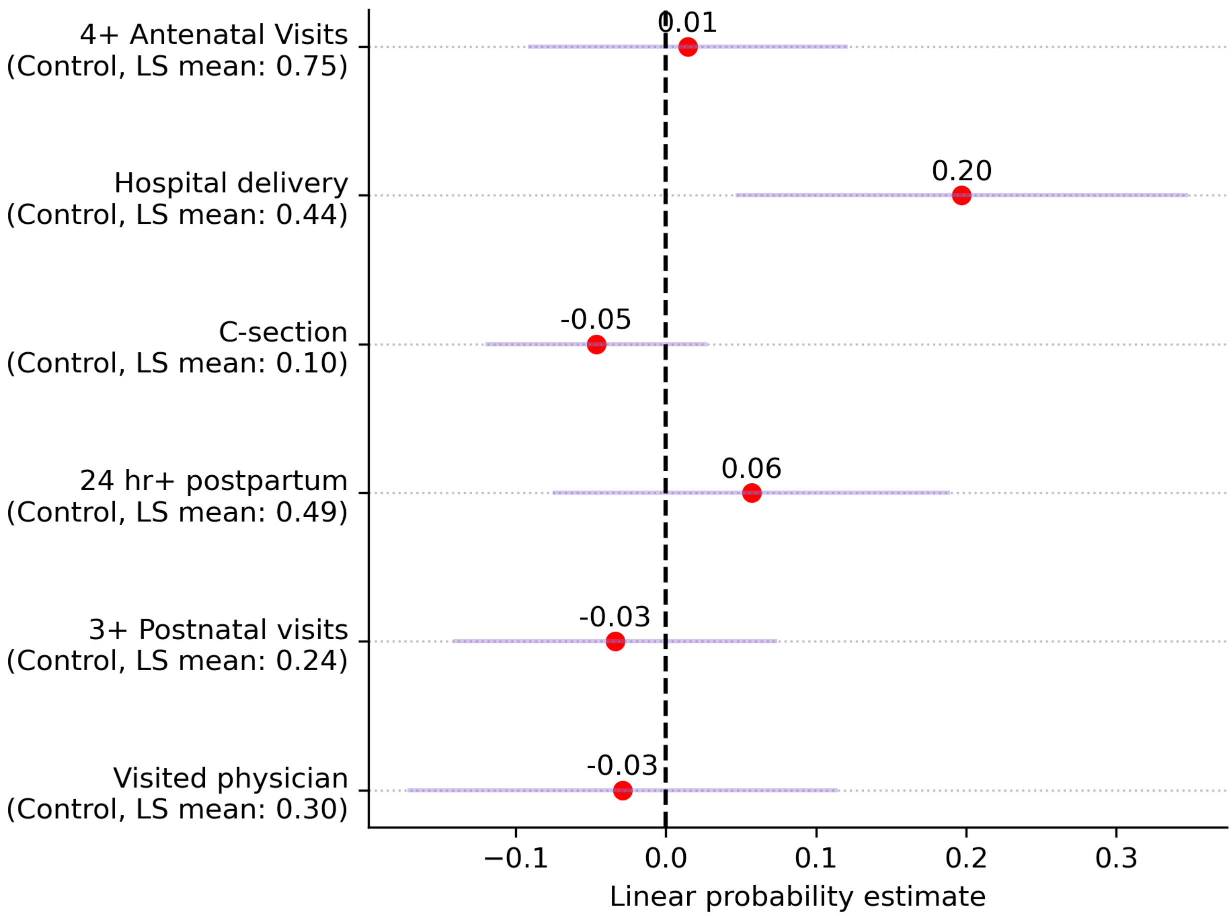
Panel A: Fertility by Period Relative to UCT Start



Panel B: IMR by Period Relative to UCT Start

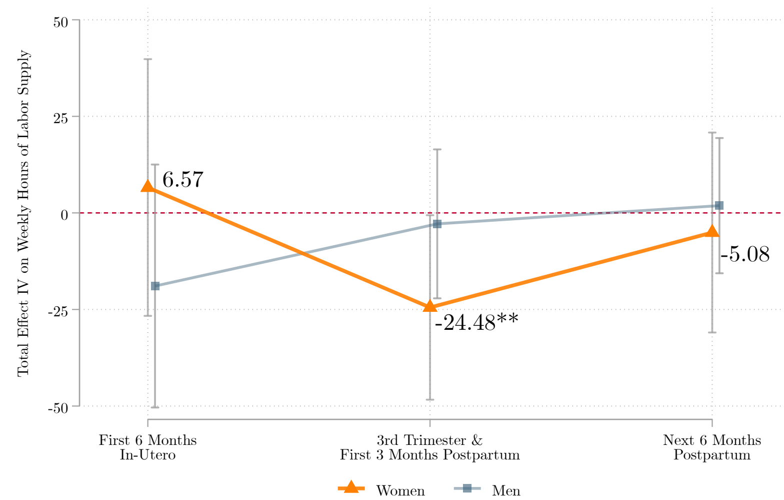
Notes: This figure is based on the sample of transfer-eligible women from the EL3 census living in households present at baseline (Panel A) and the main EL3 birth census sample, which comprises births in transfer-eligible households present at baseline (Panel B). All treatment effects are estimated using Equation (3.1). Panel A reports treatment effects in high-saturation areas on the number of births per 100 women in a given period, with birth rates standardized to a nine-month window to facilitate comparisons. Panel B reports treatment effects in high-saturation areas on infant mortality in a given period. The whiskers on each estimate denote the 95% confidence interval. Standard errors are clustered at the sublocation level.

Figure 3.6: Unconditional Cash Transfers and Healthcare Utilization, 2015-17

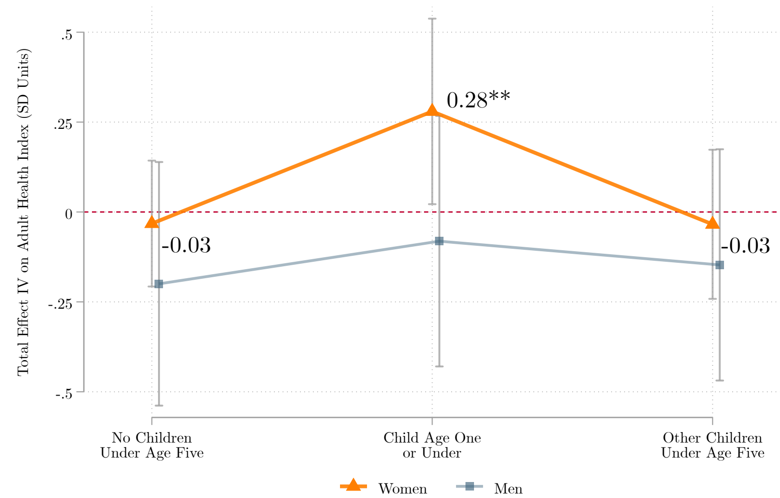


Notes: This figure is based on the sample of transfer-eligible households present at baseline and surveyed in the EL3 long-form survey. It reports treatment effects on indicators for visiting the indicated health service during a pregnancy associated with a birth between 2015 and 2017. Treatment effects are estimated using Equation (3.2). 95% confidence intervals are constructed using spatial HAC standard errors with a 10km cutoff (Conley, 2008). LS refers to low saturation sublocations. The maximum radius is fixed to 2km to match the value selected in Table 3.1. Estimates use survey data from eligible households (N=1,154 births). * $p < .10$, ** $p < .05$, *** $p < .01$.

Figure 3.7: Transfer Effects on Household Labor Supply and Health by Gender



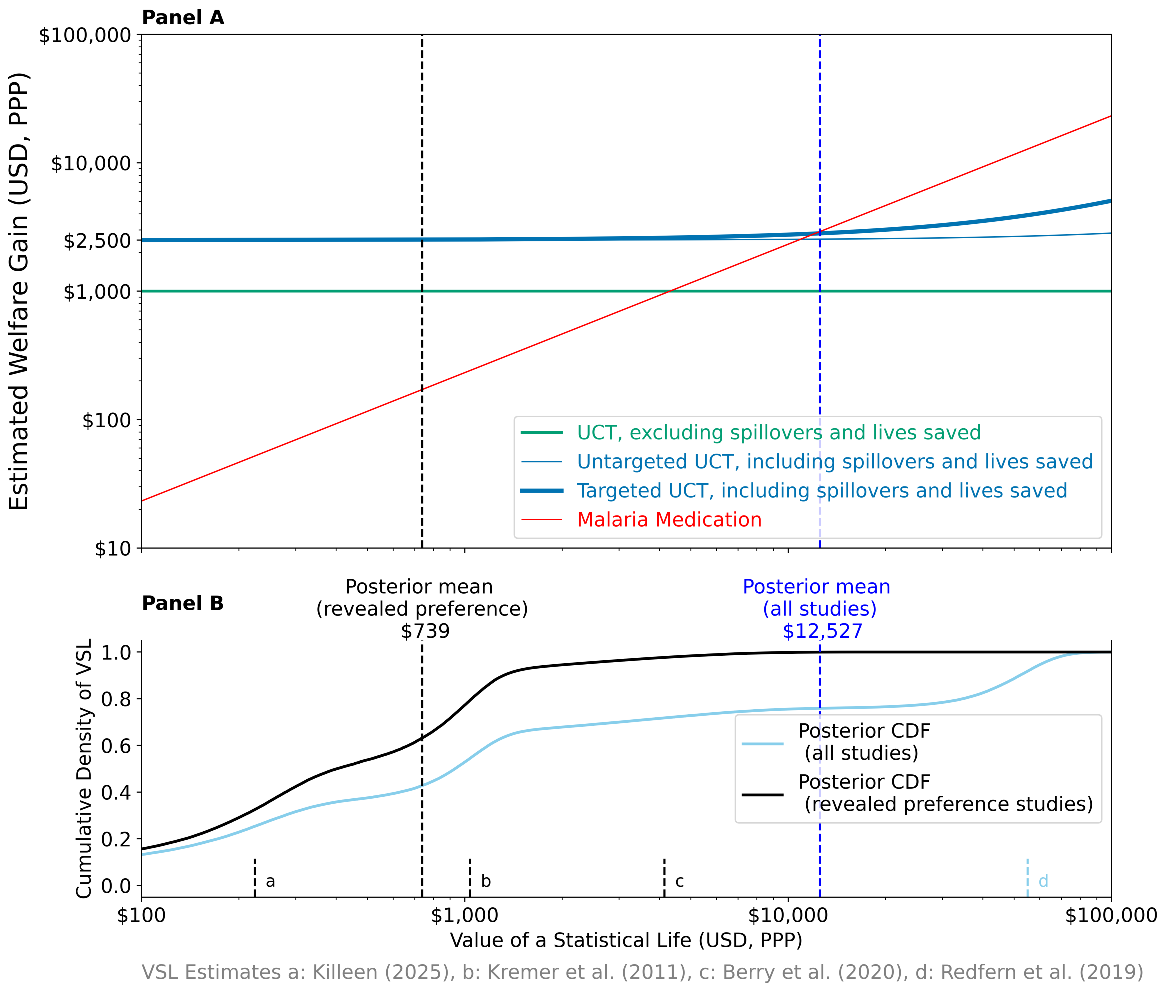
Panel A: UCTs and Labor Supply by Gender Around the Time of Birth



Panel B: UCTs and Adult Health by Gender and Presence of Recent Birth

Notes: Panel A is based on households surveyed in the first endline survey (2016-17) with a recorded birth in the EL3 census. The labor supply measures includes hours worked across the household (in agricultural employment, non-agricultural self-employment, and wage employment), as well as hours spent searching for work, in the week prior to the first endline survey. Panel A displays the estimated differential effect of the cash transfers on labor supply for three groups: households with a woman in the first 6 months of pregnancy when surveyed, households with a woman in the third trimester of pregnancy or who gave birth in the past 3 months when surveyed, and households with a woman who gave birth 3-9 months ago when surveyed. Treatment effects are estimated separately for each group using Equation (3.2), augmented with interactions with the group of interest. The control means are 45.85 (women) and 34.77 (men). N=1766 (women) and N=1677 (men). The results shown in this figure are additionally presented in Appendix Table C.12. Panel B is based on households surveyed in the first endline survey. The outcome of interest is a pre-specified, standardized index of adult health. Panel B displays the estimated differential effect of transfers on adult health for three groups: households of reproductive age with no children under age five, households with a child age one or under, and households with a child between age one and five. Spatial HAC standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$.

Figure 3.8: Welfare Gains from UCTs or Malaria Medicine by Value of a Statistical Life



Notes: Panel A of this figure plots the estimated welfare gains from a $1,000 investment in UCTs or GiveWell's top recommended program (as of March 2025), malaria medicine, varying as a function of the value of a statistical life (VSL) on the horizontal axis. We consider three UCT scenarios. First, a UCT excluding spillovers and lives saved (in which $1,000 of spending generates $1,000 of benefits), the green line. Second, we plot benefits from a saturated or at-scale UCT assuming the multiplier of 2.5 reported in Egger, Haushofer, Edward Miguel, Niehaus, et al. (2022b), and including the benefits of the child mortality reductions estimated in this paper, obtained by multiplying the VSL by 1,000 over the cost per life saved (the thin blue line). Third, We consider a targeted transfer to women in the third trimester of pregnancy, with the same spillover effects from Egger, Haushofer, Edward Miguel, Niehaus, et al., 2022b and child mortality benefits estimated in this paper (the thick blue line). Malaria medicine benefits are estimated using the cost per life saved reported by GiveWell of $4,304 ("GiveWell directed grants to top charities with impact information (2020 onward)," `https://www.givewell.org/impact-estimates`, accessed June 2025), the red line.

Panel B reports a posterior distribution of the VSL estimates in this sample obtained from Killeen (2025), Kremer, Leino, et al. (2011), Berry, Fischer, and Guiteras (2020), and Redfern et al. (2019) using Bayesian hierarchical meta analysis with a log-normal prior. The Redfern et al., 2019 estimate is not obtained via revealed preference, so we also report a revealed preference studies cumulative density function.

Table 3.1: Unconditional Cash Transfers and Mortality, 2015-17

	Reduced-Form		Spatial IV	
	(1) Infant Mortality	(2) Child Mortality	(3) Infant Mortality	(4) Child Mortality
Own village	-5.74	-11.96*	-7.98*	-12.72**
	(5.85)	(6.38)	(4.82)	(5.55)
MHT adjusted p-value	[0.234]	[0.110]		
High-saturation spillovers	-12.13**	-5.68	-11.49*	-12.91
	(5.04)	(6.66)	(6.84)	(8.12)
ATE in high-saturation sublocations	-17.87***	-17.64***	-19.46***	-25.63***
	(4.94)	(5.86)	(6.94)	(8.54)
MHT adjusted p-value			[0.044]	[0.036]
Percent reduction in HS sublocations	44.44%	30.75%	48.40%	44.67%
Control Mean	40.21	57.37	40.21	57.37
Observations	6,317	6,318	6,317	6,318

Notes: This table is based on the main EL3 birth census sample, which encompasses births from 2015-17 to transfer-eligible households present at baseline. Infant and child mortality estimated effects are reported per 1,000 live births. The ATE in high-intensity villages equals the average total effect of own-village estimates and spillovers in high-saturation sublocations. Columns (1) - (2) report estimates from equation (3.1) and columns (3) - (4) report estimation from equation (3.2). MHT corrected p-values in brackets for outcomes that were pre-specified calculated using a Romano and Wolf (2005) step-down correction based on randomization inference with 500 iterations. Reduced form standard errors are clustered at the sublocation level. Spatial HAC standard errors (Conley, 2008) with a cutoff of 10km are reported for IV estimates. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table 3.2: Heterogeneity: Complementarity with Health Services (Infant Mortality)

	Delivery in facility type (Survey data)		Infant mortality (Census data)	
	(1)	(2)	(3)	(4)
	Hospital	Physician-staffed facility	Hospital	Physician-staffed facility
Panel A: Spatial IV, no controls				
Total Effect IV	0.11	-0.04	-19.99**	-21.12***
(Time: Above median)	(0.09)	(0.07)	(8.41)	(7.75)
Total Effect x Below median time to facility	0.27**	0.07	-1.03	-2.28
	(0.14)	(0.12)	(15.28)	(15.55)
Panel B: Spatial IV, double-partial out LASSO controls + treatment interactions				
Total Effect x Below median time to facility	0.33**	0.04	-12.70	-29.01*
	(0.13)	(0.14)	(15.73)	(16.16)
Mean time to facility (above median, minutes)	52.6	55.5	52.1	54.8
Mean time to facility (below median, minutes)	18.8	22.5	19.1	22.6
Control Mean, > med. time	0.39	0.26	37.92	35.95
Control Mean, ≤ med. time	0.45	0.30	43.17	45.71
Observations	1,110	1,075	6,311	6,311

Notes: This table uses data from both the EL3 survey (columns 1-2) and the EL3 census sample (columns 3-4) for births between 2015-2017. In column (1), the outcome is birth in a hospital, based on below vs above-median travel to a hospital. Column (2) considers is similar but based on delivery in and time to a physician-staffed facility. Columns (3) and (4) consider an indicator for infant mortality, scaled to be reported in deaths per 1,000 live births, based on time to a hospital and physician-staffed facility respectively. Column (1) and (3) includes level 4 and higher facilities surveyed, plus those categorized level 4 or higher on the Kenya Master Health Facility List that were not surveyed. If the facility reported it was open in 2014 and had a physician employed during the survey the facility is included in columns (2) and (4). Rows under "Double-partial out LASSO controls, with treatment interactions" include covariates selected by double-partial out LASSO. The possible covariates includes malaria suitability, rainfall, baseline village income and assets, proximity to a road, population, distance to a town, and proximity to a water source. Covariate times treatment interactions are also included. Spatial HAC standard errors in parentheses.
* $p < .10$, ** $p < .05$, *** $p < .01$.

Bibliography

"A Practical Guide to Random-Effects Bayesian Meta-Analyses With Application to the Psychological Trauma and Suicide Literature" (July 2022). In: *Psychological trauma : theory, research, practice and policy* 15 (1), p. 121. ISSN: 1942969X. DOI: 10.1037/TRA0001316. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC10021079/.

Abadie, Alberto, Jiaying Gu, and Shu Shen (Mar. 2023). "Instrumental variable estimation with first-stage heterogeneity". In: *Journal of Econometrics*, p. 105425. ISSN: 0304-4076. DOI: 10.1016/J.JECONOM.2023.02.005.

Achuka, Vincent and Nyambega Gisesa (2020). *Uhuru Declares Curfew in War on Coronavirus*. https://nation.africa/kenya/news/uhuru-declares-curfew-in-war-on-coronavirus-281868. Accessed: 2025-04-12.

Agness, Daniel et al. (Jan. 2025). "Valuing the Time of the Self-Employed". In: *The Review of Economic Studies*, rdaf003. ISSN: 0034-6527. DOI: 10.1093/restud/rdaf003. eprint: https://academic.oup.com/restud/advance-article-pdf/doi/10.1093/restud/rdaf003/61458569/rdaf003.pdf. URL: https://doi.org/10.1093/restud/rdaf003.

Agrawal, Virat, Neeraj Sood, and Christopher M. Whaley (Oct. 2023). "The Impact of the Global COVID-19 Vaccination Campaign on All-Cause Mortality". In: *NBER Working Paper*. DOI: 10.3386/W31812. URL: https://www.nber.org/papers/w31812.

Ahuja, Amrita et al. (2021). "Preparing for a Pandemic: Accelerating Vaccine Availability". In: *AEA Papers and Proceedings* 111, pp. 331–35. ISSN: 2574-0768. DOI: 10.1257/PANDP.20211103. URL: https://doi.org/10.1257/pandp.20211103.

Aldy, Joseph E. and W. Kip Viscusi (July 2007). "Age Differences in the Value of Statistical Life: Revealed Preference Evidence". In: *https://doi.org/10.1093/reep/rem014* 1.2, pp. 241–260. ISSN: 1750-6816. DOI: 10.1093/REEP/REM014. URL: https://www.journals.uchicago.edu/doi/10.1093/reep/rem014.

Amarante, Verónica et al. (2016). "Do Cash Transfers Improve Birth Outcomes? Evidence from Matched Vital Statistics, Program, and Social Security Data". In: *American Economic Journal: Economic Policy* 8.2, pp. 1–43.

Amek, Nyaguara O., Frank O. Odhiambo, et al. (Dec. 2014). "Childhood cause-specific mortality in rural Western Kenya: application of the InterVA-4 model". In: *Global Health Action* 7.1. Publisher: Taylor & Francis _eprint: https://doi.org/10.3402/gha.v7.25581, p. 25581. ISSN: 1654-9716. DOI: 10.3402/gha.v7.25581. URL: https://doi.org/10.3402/gha.v7.25581 (visited on 11/28/2023).

Amek, Nyaguara O., Annemieke Van Eijk, et al. (Jan. 2018). "Infant and child mortality in relation to malaria transmission in KEMRI/CDC HDSS, Western Kenya: validation of verbal autopsy". In: *Malaria Journal* 17.1, p. 37. ISSN: 1475-2875. DOI: 10.1186/s12936-018-2184-x. URL: https://doi.org/10.1186/s12936-018-2184-x (visited on 11/28/2023).

Anderson, T W and Herman Rubin (1949). "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations". In: *Annals of Mathematical Statistics* 20.1.

Andrew, Alison and Marcos Vera-Hernández (Jan. 2024). "Incentivizing Demand for Supply-Constrained Care: Institutional Birth in India". In: *The Review of Economics and Statistics* 106.1, pp. 102–118. ISSN: 0034-6535. DOI: 10.1162/rest_a_01206. eprint: https://direct.mit.edu/rest/article-pdf/106/1/102/2354813/rest_a_01206.pdf. URL: https://doi.org/10.1162/rest%5C_a%5C_01206.

Ashenfelter, Orley (Mar. 2006). "Measuring the Value of a Statistical Life: Problems and Prospects*". In: *The Economic Journal* 116.510, pp. C10–C23. ISSN: 1468-0297. DOI: 10.1111/J.1468-0297.2006.01072.X. URL: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1468-0297.2006.01072.x%20https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0297.2006.01072.x%20https://onlinelibrary.wiley.com/doi/10.1111/j.1468-0297.2006.01072.x.

Ashenfelter, Orley and Michael Greenstone (2004). "Estimating the Value of a Statistical Life: The Importance of Omitted Variables and Publication Bias". In: *American Economic Review* 94.2. ISSN: 0002-8282. DOI: 10.1257/0002828041301984.

Ashraf, Nava, James Berry, and Jesse M. Shapiro (2010). "Can Higher Prices Stimulate Product Use? Evidence from a Field Experiment in Zambia". In: *American Economic Review* 100.5. ISSN: 0002-8282. DOI: 10.1257/AER.100.5.2383. URL: http://www.aeaweb.org/articles.php?doi=10.1257/aer.100.5.2383.

Ashraf, Nava, Dean Karlan, and Wesley Yin (May 2006). "Tying Odysseus to the Mast: Evidence From a Commitment Savings Product in the Philippines*". In: *The Quarterly Journal of Economics* 121.2, pp. 635–672. ISSN: 0033-5533. DOI: 10.1162/qjec.2006.121.2.635. eprint: https://academic.oup.com/qje/article-pdf/121/2/635/5324429/121-2-635.pdf. URL: https://doi.org/10.1162/qjec.2006.121.2.635.

Asker, Erdal, Shatakshee Dhongde, and Abu S. Shonchoy (2025). "COVID-19 and mortality among infants: Evidence from India". In: *Journal of Health Economics* 101, p. 102991. ISSN: 0167-6296. DOI: https://doi.org/10.1016/j.jhealeco.2025.102991. URL: https://www.sciencedirect.com/science/article/pii/S0167629625000256.

Atkin, David et al. (Aug. 2017). "Organizational Barriers to Technology Adoption: Evidence from Soccer-Ball Producers in Pakistan". In: *The Quarterly Journal of Economics* 132.3, pp. 1101–1164. ISSN: 0033-5533. DOI: 10.1093/QJE/QJX010. URL: https://academic.oup.com/qje/article/132/3/1101/3064351.

Babagoli, Masih A. et al. (Sept. 2022). "Cost-Effectiveness and Cost–Benefit Analyses of Providing Menstrual Cups and Sanitary Pads to Schoolgirls in Rural Kenya". In: *Women's Health Reports* 3.1, p. 773. DOI: 10.1089/WHR.2021.0131. URL: /pmc/articles/

PMC9518800/%20/pmc/articles/PMC9518800/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9518800/.

Bachani, A. M. et al. (Mar. 2017). "Helmet wearing in Kenya: prevalence, knowledge, attitude, practice and implications". In: *Public Health* 144, S23–S31. ISSN: 0033-3506. DOI: 10.1016/J.PUHE.2016.12.005.

Bai, Jie, David Sungho Park, and Ajay Shenoy (2025). "Misperceptions and Product Choice: Evidence from a Randomized Trial in Zambia *". In.

Baird, Sarah, Craig McIntosh, and Berk Özler (Sept. 2019). "When the money runs out: Do cash transfers have sustained effects on human capital accumulation?" en. In: *Journal of Development Economics* 140, pp. 169–185. ISSN: 0304-3878. DOI: 10.1016/j.jdeveco.2019.04.004. URL: https://www.sciencedirect.com/science/article/pii/S0304387818312732 (visited on 06/03/2021).

Banerjee, Abhijit Vinayak et al. (June 2010). "Improving immunisation coverage in rural India: clustered randomised controlled evaluation of immunisation campaigns with and without incentives". In: *BMJ (Clinical research ed.)* 340.7759, p. 1291. ISSN: 1756-1833. DOI: 10.1136/BMJ.C2220. URL: https://pubmed.ncbi.nlm.nih.gov/20478960/.

Bartel, Ann et al. (2023). "The Impacts of Paid Family and Medical Leave on Worker Health, Family Well-Being, and Employer Outcomes". In: *Annual Review of Public Health* 44.Volume 44, 2023, pp. 429–443. ISSN: 1545-2093. DOI: https://doi.org/10.1146/annurev-publhealth-071521-025257. URL: https://www.annualreviews.org/content/journals/10.1146/annurev-publhealth-071521-025257.

Bassi, Vittorio et al. (Nov. 2022). "Achieving Scale Collectively". In: *Econometrica* 90.6, pp. 2937–2978. ISSN: 1468-0262. DOI: 10.3982/ECTA18773. URL: https://onlinelibrary.wiley.com/doi/full/10.3982/ECTA18773%20https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA18773%20https://onlinelibrary.wiley.com/doi/10.3982/ECTA18773.

Bastagli, Francesca, Jessica Hagen-Zanker, and Georgina Sturge (July 2016). *Cash transfers: what does the evidence say?* URL: https://odi.org/en/publications/cash-transfers-what-does-the-evidence-say-a-rigorous-review-of-impacts-and-the-role-of-design-and-implementation-features/ (visited on 11/28/2023).

Baylis, Patrick et al. (2023). "Is the Demand for Clean Air Too Low? Experimental Evidence from Delhi". URL: https://michaelgreenstone.com/wp-content/uploads/2024/04/sahai_india_wtp.pdf.

BBC (2017). *Kenya's Uhuru Kenyatta Declares Drought a National Disaster*. Accessed: 2025-02-23. URL: https://www.bbc.com/news/world-africa-38934847.

Becker, Gordon M., Morris H. Degroot, and Jacob Marschak (1964). "Measuring utility by a single-response sequential method". In: *Behavioral Science* 9.3, pp. 226–232. ISSN: 00057940. DOI: 10.1002/bs.3830090304. URL: http://doi.wiley.com/10.1002/bs.3830090304.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen (2014a). "Inference on Treatment Effects after Selection among High-Dimensional Controls". In: *The Review* *of Economic Studies* 81.2. ISSN: 0034-6527. DOI: 10.1093/RESTUD/RDT044. URL: https://academic.oup.com/restud/article/81/2/608/1523757.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen (2014b). "Inference on Treatment Effects after Selection among High-Dimensional Controls". In: *The Review of Economic Studies* 81.2, pp. 608–650. DOI: 10.1093/restud/rdt044. URL: https://doi.org/10.1093/restud/rdt044.

Benjamin, Daniel J, Don A Moore, and Matthew Rabin (Oct. 2017). "Biased Beliefs About Random Samples: Evidence from Two Integrated Experiments". In: *NBER Working Paper*. DOI: 10.3386/W23927. URL: https://www.nber.org/papers/w23927.

Bergquist, Lauren Falcao and Michael Dinerstein (Dec. 2020). "Competition and Entry in Agricultural Markets: Experimental Evidence from Kenya". In: *American Economic Review* 110.12, pp. 3705–47. ISSN: 0002-8282. DOI: 10.1257/AER.20171397. URL: https://doi.org/10.1257/aer.20171397.

Berry, James, Greg Fischer, and Raymond Guiteras (2020). "Eliciting and Utilizing Willingness to Pay: Evidence from Field Trials in Northern Ghana". In: *Journal of Political Economy* 128.4. ISSN: 1537534X. DOI: 10.1086/705374. URL: https://www.journals.uchicago.edu/doi/10.1086/705374.

Blattman, Christopher and Paul Niehaus (2014). "Show Them the Money: Why Giving Cash Helps Alleviate Poverty". In: *Foreign Affairs* 93.3. Accessed: 2025-02-23, pp. 117–126. URL: http://www.jstor.org/stable/24483411.

Bloom, Nicholas, Benn Eifert, et al. (Feb. 2013). "Does Management Matter? Evidence from India *". In: *The Quarterly Journal of Economics* 128.1, pp. 1–51. ISSN: 0033-5533. DOI: 10.1093/qje/qjs044. URL: https://doi.org/10.1093/qje/qjs044.

Bloom, Nicholas, Aprajit Mahajan, et al. (May 2010). "Why Do Firms in Developing Countries Have Low Productivity?" In: *American Economic Review* 100.2, pp. 619–23. ISSN: 0002-8282. DOI: 10.1257/AER.100.2.619. URL: http://www.aeaweb.org/articles.php?doi=10.1257/aer.100.2.619.

Bolton, Patrick and Christopher Harris (1999). "Strategic Experimentation". In: *Econometrica* 67.2, pp. 349–374. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/2999588.

Buera, Francisco J., Joseph P. Kaboski, and Yongseok Shin (Aug. 2011). "Finance and Development: A Tale of Two Sectors". In: *American Economic Review* 101.5, pp. 1964–2002. ISSN: 0002-8282. DOI: 10.1257/AER.101.5.1964. URL: http://www.aeaweb.org/articles.php?doi=10.1257/aer.101.5.1964.

Buffa, Giavana et al. (2018). "Prenatal stress and child development: A scoping review of research in low- and middle-income countries". In: *PLOS ONE* 13.12, e0207235. DOI: 10.1371/journal.pone.0207235. URL: https://doi.org/10.1371/journal.pone.0207235.

Burke, Marshall, Lauren Falcao Bergquist, and Edward Miguel (2019). "Sell Low and Buy High: Arbitrage and Local Price Effects in Kenyan Markets". In: *Quarterly Journal of Economics* 134.2, pp. 785–842. DOI: 10.1093/qje/qjy034.

Burstein, Roy et al. (Oct. 2019). "Mapping 123 million neonatal, infant and child deaths between 2000 and 2017". In: *Nature* 574.7778, pp. 353–358. ISSN: 1476-4687. DOI: 10.1038/s41586-019-1545-0. URL: https://doi.org/10.1038/s41586-019-1545-0.

Cai, Chenxi et al. (Mar. 2020). "The impact of occupational activities during pregnancy on pregnancy outcomes: a systematic review and metaanalysis". In: *American Journal of Obstetrics and Gynecology* 222 (3), pp. 224–238. ISSN: 10976868. DOI: 10.1016/j.ajog.2019.08.059. URL: https://pubmed.ncbi.nlm.nih.gov/31550447/.

Callaway, Brantly and Pedro H. C. Sant'Anna (Dec. 2020). "Difference-in-Differences with multiple time periods". en. In: *Journal of Econometrics*. ISSN: 0304-4076. DOI: 10.1016/j.jeconom.2020.12.001. URL: https://www.sciencedirect.com/science/article/pii/S0304407620303948 (visited on 05/28/2021).

Campbell, Harry F. and Richard P. C. Brown (July 2003). "Valuation of Non-marketed Goods". In: *Benefit-Cost Analysis*, pp. 261–287. DOI: 10.1017/CBO9780511791291.013. URL: https://www.cambridge.org/core/books/benefitcost-analysis/valuation-of-nonmarketed-goods/DB15304FB7C68AA09F29D79E70CC7FB4.

Carneiro, Pedro et al. (Aug. 2021). *Do Cash Transfers Promoting Early Childhood Development Have Unintended Effects on Fertility? Evidence from Northern Nigeria*. Tech. rep. URL: https://www.povertyactionlab.org/sites/default/files/research-paper/working_paper_914_Do-Cash-Transfers-Have-Unintended-Effects-On-Fertility_Nigeria_Aug2021.pdf.

Casaburi, Lorenzo and Jack Willis (Dec. 2018). "Time versus State in Insurance: Experimental Evidence from Contract Farming in Kenya". In: *American Economic Review* 108.12, pp. 3778–3813. ISSN: 0002-8282. DOI: 10.1257/AER.20171526.

Chatterjee, Shoumitro and Tom Vogl (2018). "Escaping Malthus: Economic Growth and Fertility Change in the Developing World". In: *American Economic Review* 108.6, pp. 1440–67. DOI: 10.1257/aer.20170748. URL: https://www.aeaweb.org/articles?id=10.1257/aer.20170748.

Cirera, Xavier, Diego Comin, and Marcio Cruz (2022). "Bridging the Technological Divide Technology Adoption by Firms in Developing Countries". In: *World Bank Productivity Project*.

Cohen, Jessica and Pascaline Dupas (2010). "Free distribution or cost-sharing? Evidence from a randomized malaria prevention experiment". In: *Quarterly Journal of Economics* 125.1, pp. 1–45. ISSN: 00335533. DOI: 10.1162/QJEC.2010.125.1.1/2/125-1-1.PDF.GIF. URL: https://dx.doi.org/10.1162/qjec.2010.125.1.1.

Colmer, Jonathan (Sept. 2020). "What is the meaning of (statistical) life? Benefit–cost analysis in the time of COVID-19". In: *Oxford Review of Economic Policy* 36.Supplement_1, S56–S63. ISSN: 0266-903X. DOI: 10.1093/OXREP/GRAA022. URL: https://dx.doi.org/10.1093/oxrep/graa022.

Comin, Diego and Martí Mestieri (July 2018). "If Technology Has Arrived Everywhere, Why Has Income Diverged?" In: *American Economic Journal: Macroeconomics* 10.3, pp. 137–78. ISSN: 1945-7707. DOI: 10.1257/MAC.20150175. URL: https://doi.org/10.1257/mac.20150175.

Conley, Timothy G. (2008). "Spatial Econometrics". In: *The New Palgrave Dictionary of Economics*. Ed. by Steven N. Durlauf and Lawrence E. Blume. Second Edition. Vol. 7. Houndsmills: Palgrave Macmillan, pp. 741–47.

Crosta, Tommaso et al. (2024). *Unconditional Cash Transfers: A Bayesian Meta-Analysis of Randomized Evaluations in Low and Middle Income Countries*. Working Paper 32779. National Bureau of Economic Research. DOI: 10.3386/w32779. URL: http://www.nber.org/papers/w32779.

Cutler, David, Adriana Lleras-Muney, and Tom Vogl (2012). "Socioeconomic Status and Health: Dimensions and Mechanisms". In: *The Oxford Handbook of Health Economics*. New York: Oxford University Press.

Deaton, Angus (2016). "Measuring and Understanding Behavior, Welfare, and Poverty". In: *American Economic Review* 106.6, pp. 1221–43. DOI: 10.1257/aer.106.6.1221. URL: https://www.aeaweb.org/articles?id=10.1257/aer.106.6.1221.

Deaton, Angus S. and Christina Paxson (2004). "Mortality, Income, and Income Inequality over Time in Britain and the United States". In: *Perspectives on the Economics of Aging*. Ed. by David A. Wise. University of Chicago Press, pp. 247–286.

Doraszelski, Ulrich, Gregory Lewis, and Ariel Pakes (Mar. 2018). "Just Starting Out: Learning and Equilibrium in a New Market". In: *American Economic Review* 108.3, pp. 565–615. ISSN: 0002-8282. DOI: 10.1257/AER.20160177.

Dupas, P. and E. Miguel (Jan. 2017). "Impacts and Determinants of Health Levels in Low-Income Countries". In: 2, pp. 3–93. ISSN: 2214-658X. DOI: 10.1016/BS.HEFE.2016.09.003.

Dupas, Pascaline and Jonathan Robinson (2013). "Why Don't the Poor Save More? Evidence from Health Savings Experiments". In: *American Economic Review* 103.4, pp. 1138–71. DOI: 10.1257/aer.103.4.1138. URL: https://www.aeaweb.org/articles?id=10.1257/aer.103.4.1138.

Egger, Dennis, Johannes Haushofer, Edward Miguel, Paul Niehaus, et al. (Nov. 2022a). "General Equilibrium Effects of Cash Transfers: Experimental Evidence From Kenya". In: *Econometrica* 90.6, pp. 2603–2643. ISSN: 1468-0262. DOI: 10.3982/ECTA17945.

— (2022b). "General Equilibrium Effects of Cash Transfers: Experimental Evidence From Kenya". In: *Econometrica* 90.6, pp. 2603–2643. ISSN: 1468-0262. DOI: 10.3982/ECTA17945. (Visited on 02/14/2023).

Egger, Dennis, Johannes Haushofer, Edward Miguel, and Michael Walker (2023). *General Equilibrium Effects of Cash Transfers: Pre-analysis plan for Endline 3 (EL3) Child Mortality Analysis*. Pre-analysis plan, posted on the AEA Social Science Registry. URL: https://www.socialscienceregistry.org/versions/250281/docs/version/document.

Egger, Dennis, Grady Killeen, et al. (2024). *Amendment to: General Equilibrium Effects of Cash Transfers: Pre-analysis plan for Endline 3 (EL3) Child Mortality Analysis*. Amendment to pre-analysis plan, AEA Social Science Registry. URL: https://www.socialscienceregistry.org/versions/250293/docs/version/document.

Evans, David and Katrina Kosec (2016). *Cash Transfers and Health: It Matters When You Measure and It Matters How Many Health Care Workers*. https://blogs.worldbank.org/en/impactevaluations/cash-transfers-and-health-it-matters-when-you-measure-and-it-matters-how-many-health-care-workers. Accessed: 2025-02-23.

Evans, Jonathan et al. (May 2025). *Spiritual and Religious Beliefs and Practices Around the World — Pew Research Center*. URL: https://www.pewresearch.org/religion/2025/05/06/believing-in-spirits-and-life-after-death-is-common-around-the-world/.

Fafchamps, Marcel et al. (Jan. 2014). "Microenterprise growth and the flypaper effect: Evidence from a randomized experiment in Ghana". In: *Journal of Development Economics* 106, pp. 211–226. ISSN: 0304-3878. DOI: 10.1016/J.JDEVECO.2013.09.010. URL: https://www.sciencedirect.com/science/article/pii/S0304387813001375.

Foster, A. D. and M. R. Rosenzweig (1995). "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture". In: *https://doi.org/10.1086/601447* 103.6, pp. 1176–1209. ISSN: 00223808. DOI: 10.1086/601447. URL: https://www.journals.uchicago.edu/doi/10.1086/601447.

Gacheri, Susan et al. (2014). "Performance of clinicians in identifying tuberculosis as cause of death using verbal autopsy questionnaires in Siaya County, Kenya". In: *African Journal of Health Sciences* 27, pp. 232–238.

Garganta, Santiago, Leonardo Gasparini, and Mariana Marchionni (2017). "Cash Transfers and Female Labor Force Participation: The Case of AUH in Argentina". In: *IZA Journal of Labor Policy* 6, pp. 1–22.

Gertler, Paul and Jonathan Gruber (Mar. 2002a). "Insuring Consumption Against Illness". In: *American Economic Review* 92.1, pp. 51–70.

— (2002b). "Insuring Consumption Against Illness". In: *American Economic Review* 92.1, pp. 51–70. URL: https://www.aeaweb.org/articles?id=10.1257/000282802760015603.

Gillen, Ben, Erik Snowberg, and Leeat Yariv (2019). "Experimenting with measurement error: Techniques with applications to the caltech cohort study". In: *Journal of Political Economy* 127.4. ISSN: 1537534X. DOI: 10.1086/701681/SUPPL{_}FILE/2016930DATA.ZIP. URL: https://www.journals.uchicago.edu/doi/10.1086/701681.

Greenberg, Kyle et al. (2021). "The Heterogeneous Value of a Statistical Life: Evidence from U.S. Army Reenlistment Decisions". In: *NBER Working Paper* No. 29104. DOI: 10.3386/w29104. URL: http://www.nber.org/papers/w29104%20http://www.nber.org/papers/w29104.pdf.

Greenstone, Michael and B. Kelsey Jack (Mar. 2015). "Envirodevonomics: A Research Agenda for an Emerging Field". In: *Journal of Economic Literature* 53.1, pp. 5–42. ISSN: 0022-0515. DOI: 10.1257/JEL.53.1.5. URL: http://dx.doi.org/10.1257/jel.53.1.5.

Grépin, Karen A., James Habyarimana, and William Jack (2019). "Cash on delivery: Results of a randomized experiment to promote maternal health care in Kenya". In: *Journal of Health Economics* 65, pp. 15–30. DOI: \url{https://doi.org/10.1016/j.jhealeco.2018.12.001.}.

Guldi, Melanie et al. (2024). "Supplemental Security Income for Children, Maternal Labor Supply, and Family Well-Being: Evidence from Birth Weight Eligibility Cutoffs". In: *Journal of Human Resources* 59.4, pp. 975–1010.

Hall, Robert E. and Charles I. Jones (Feb. 2007). "The Value of Life and the Rise in Health Spending". In: *The Quarterly Journal of Economics* 122.1, pp. 39–72. ISSN: 0033-5533. DOI: 10.1162/QJEC.122.1.39. URL: https://academic.oup.com/qje/article/122/1/39/1924761.

Hamze, Hasan, Asrat Mengiste, and Jane 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." In: *The Pan African medical journal* 28, p. 35. ISSN: 1937-8688. DOI: 10.11604/pamj.2017.28.35.10344.

Haushofer, Johannes et al. (2025). "Targeting Impact versus Deprivation". In: *American Economic Review* 115.6, pp. 1936–74. DOI: 10.1257/aer.20221650. URL: https://www.aeaweb.org/articles?id=10.1257/aer.20221650.

Havránek, Tomáš (2015). "Measuring Intertemporal Substitution: The Importance of Method Choices and Selective Reporting". In: *Journal of the European Economic Association* 13.6. ISSN: 1542-4766. DOI: 10.1111/JEEA.12133. URL: https://dx.doi.org/10.1111/jeea.12133.

Heath, Rachel and Seema Jayachandran (2018). "The Causes and Consequences of Increased Female Education and Labor Force Participation in Developing Countries". In: *The Oxford Handbook of Women and the Economy*. Ed. by Susan L. Averett, Laura M. Argys, and Saul D. Hoffman. Online edition published July 6, 2017. Accessed 27 July 2025. Oxford University Press. DOI: 10.1093/oxfordhb/9780190628963.013.10. URL: https://doi.org/10.1093/oxfordhb/9780190628963.013.10.

Institute for Health Metrics and Evaluation (2014). *Health Service Provision in Kenya: Assessing Facility Capacity, Costs of Care, and Patient Perspectives*. Seattle, WA: Institute for Health Metrics and Evaluation (IHME).

— (2025). *Verbal Autopsy Tool*. https://www.healthdata.org/data-tools-practices/verbal-autopsy. Accessed: 2025-02-22.

International Labour Organization (2019). *Small Matters: Global evidence on the contribution to employment by the self-employed, micro-enterprises and SMEs*. Tech. rep. United Nations. URL: https://www.ilo.org/publications/small-matters-global-evidence-contribution-employment-self-employed-micro.

— (2025). *Labor force participation rate [dataset]*. https://ourworldindata.org/. Processed by Our World in Data; original data from ILOSTAT via the World Bank's World Development Indicators.

Ito, Koichiro and Shuang Zhang (May 2020). "Willingness to pay for clean air: Evidence from air purifier markets in China". In: *Journal of Political Economy* 128.5, pp. 1627–1672. ISSN: 1537534X. DOI: 10.1086/705554/SUPPL{_}FILE/2016459DATA.ZIP. URL: https://www.journals.uchicago.edu/doi/full/10.1086/705554.

Izugbara, Chimaraoke O. and David P. Ngilangwa (2010). "Women, poverty and adverse maternal outcomes in Nairobi, Kenya". In: *BMC Women's Health* 10, p. 33. DOI: 10.

1186/1472-6874-10-33. URL: https://bmcwomenshealth.biomedcentral.com/articles/10.1186/1472-6874-10-33.

Jones, Charles I. and Peter J. Klenow (2016). “Beyond GDP? Welfare across Countries and Time”. In: *American Economic Review* 106.9, pp. 2426–57. DOI: 10.1257/aer.20110236. URL: https://www.aeaweb.org/articles?id=10.1257/aer.20110236.

Kader, Manzur et al. (July 2021). “Shift and night work during pregnancy and preterm birth—a cohort study of Swedish health care employees”. In: *International Journal of Epidemiology* 50.6, pp. 1864–1874. ISSN: 0300-5771. DOI: 10.1093/ije/dyab135. eprint: https://academic.oup.com/ije/article-pdf/50/6/1864/42092240/dyab135.pdf. URL: https://doi.org/10.1093/ije/dyab135.

Kansikas, Carolina, Anandi Mani, and Paul Niehaus (2025). *Structuring cash transfers: cash flow preferences, seasonality, and financial decisions in rural Kenya*. Tech. rep. UC San Diego.

Karlan, Dean et al. (May 2014). “Agricultural decisions after relaxing credit and risk constraints”. In: *Quarterly Journal of Economics* 129.2, pp. 597–652. ISSN: 15314650. DOI: 10.1093/qje/qju002. URL: https://academic.oup.com/qje/article/129/2/597/1867065.

Kenya National Bureau of Statistics (2019). *Gross County Product: 2019*.

Killeen, Grady (2025). “A New Experimental Method for Estimating Demand for Non-market Goods: With an Application to the Value of a Statistical Life”. Working paper.

Kremer, Michael, Jean Lee, et al. (May 2013). “Behavioral Biases and Firm Behavior: Evidence from Kenyan Retail Shops”. In: *American Economic Review: Papers and Proceedings* 103.3, pp. 362–68. ISSN: 0002-8282. DOI: 10.1257/AER.103.3.362. URL: http://dx.doi.org/10.1257/aer.103.3.362http://dx.doi.org/10.1257/aer.103.3.362..

Kremer, Michael, Jessica Leino, et al. (2011). “Spring Cleaning: Rural Water Impacts, Valuation, and Property Rights Institutions*”. In: *The Quarterly Journal of Economics* 126.1. ISSN: 0033-5533. DOI: 10.1093/QJE/QJQ010. URL: https://academic.oup.com/qje/article/126/1/145/1903056.

Kremer, Michael and Edward Miguel (Aug. 2007). “The Illusion of Sustainability”. In: *The Quarterly Journal of Economics* 122.3, pp. 1007–1065. ISSN: 0033-5533. DOI: 10.1162/QJEC.122.3.1007. URL: https://dx.doi.org/10.1162/qjec.122.3.1007.

Kumar, Santosh (2024). “The Effects of Childhood Immunization Program on Health and Education: Micro-Evidence from India”. URL: https://docs.iza.org/dp17083.pdf.

Lacetera, Nicola, Devin G. Pope, and Justin R. Sydnor (Aug. 2012). “Heuristic Thinking and Limited Attention in the Car Market”. In: *American Economic Review* 102.5, pp. 2206–36. ISSN: 0002-8282. DOI: 10.1257/AER.102.5.2206. URL: http://dx.doi.org/10.1257/aer.102.5.2206.

László, Krisztina D. et al. (2013). “Maternal Bereavement During Pregnancy and the Risk of Stillbirth: A Nationwide Cohort Study in Sweden”. In: *American Journal of Epidemiology* 177.3, pp. 219–227. DOI: 10.1093/aje/kws383. URL: https://doi.org/10.1093/aje/kws383.

León, Gianmarco and Edward Miguel (Jan. 2017a). "Risky Transportation Choices and the Value of a Statistical Life". In: *American Economic Journal: Applied Economics* 9.1, pp. 202–28. ISSN: 1945-7782. DOI: 10.1257/APP.20160140.

— (2017b). "Risky Transportation Choices and the Value of a Statistical Life". In: *American Economic Journal: Applied Economics* 9.1, pp. 202–28. DOI: 10.1257/app.20160140. URL: https://www.aeaweb.org/articles?id=10.1257/app.20160140.

Li, Jia and Sun Ah Kim (2022). "Manufacturer returns: An empirical study". In: *Journal of Inter-Organizational Relationships* 28.1-2, pp. 50–63. ISSN: 26943999. DOI: 10.1080/26943980.2022.2086660. URL: https://www.tandfonline.com/doi/abs/10.1080/26943980.2022.2086660.

Lima, Marcelo de (2020). "The value of a statistical life in Mexico". In: *J Environ Econ Polic* 9.2. ISSN: 21606552. DOI: 10.1080/21606544.2019.1617196. URL: https://www.tandfonline.com/doi/pdf/10.1080/21606544.2019.1617196.

Liu, Bette C. et al. (2008). "Helmets for preventing injury in motorcycle riders". In: *The Cochrane database of systematic reviews* 1. ISSN: 1469-493X. DOI: 10.1002/14651858.CD004333.PUB3. URL: https://pubmed.ncbi.nlm.nih.gov/18254047/.

Lleras-Muney, Adriana, Hannes Schwandt, and Laura Wherry (2024). *Poverty and Health*. Working Paper 32866. National Bureau of Economic Research. DOI: 10.3386/w32866. URL: http://www.nber.org/papers/w32866.

Lyons-Amos, Mark and Timothy Stones (2017). "Trends in Demographic and Health Survey data quality: an analysis of age heaping over time in 34 countries in Sub-Saharan Africa between 1987 and 2015". In: *BMC Research Notes* 10.1, p. 760. DOI: 10.1186/s13104-017-3091-x.

Ma, Lin et al. (2021). *The Intergenerational Mortality Tradeoff of COVID-19 Lockdown Policies*. Working Paper 28925. National Bureau of Economic Research. DOI: 10.3386/w28925. URL: http://www.nber.org/papers/w28925.

Manski, Charles F. (Sept. 2004). "Measuring Expectations". In: *Econometrica* 72.5, pp. 1329–1376. ISSN: 1468-0262. DOI: 10.1111/J.1468-0262.2004.00537.X. URL: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1468-0262.2004.00537.x%20https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0262.2004.00537.x%20https://onlinelibrary.wiley.com/doi/10.1111/j.1468-0262.2004.00537.x.

McIntosh, Craig and Andrew Zeitlin (2024). "Cash Versus Kind: Benchmarking a Child Nutrition Program Against Unconditional Cash Transfers in Rwanda". In: *Economic Journal* 134.664, pp. 3360–3389. DOI: 10.1093/ej/ueae050.

Meki, Muhammad (2025). "Small Firm Investment under Uncertainty: The Role of Equity Finance". URL: https://mmeki.com/wp-content/uploads/Small-Firm-Investment-under-Uncertainty.pdf.

Mel, Suresh de, David McKenzie, and Christopher Woodruff (Nov. 2008). "Returns to Capital in Microenterprises: Evidence from a Field Experiment". In: *The Quarterly Journal of Economics* 123.4, pp. 1329–1372. ISSN: 0033-5533. DOI: 10.1162/QJEC.2008.123.4.1329. URL: https://academic.oup.com/qje/article/123/4/1329/1933166.

Mel, Suresh de, David McKenzie, and Christopher Woodruff (2019). "Labor Drops: Experimental Evidence on the Return to Additional Labor in Microenterprises". In: *American Economic Journal: Applied Economics* 11.1, pp. 202–35. ISSN: 1945-7782. DOI: 10.1257/APP.20170497. URL: https://doi.org/10.1257/app.20170497.

Mendelsohn, Robert and Sheila Olmstead (Nov. 2009). "The economic valuation of environmental amenities and disamenities: Methods and applications". In: *Annual Review of Environment and Resources* 34.Volume 34, 2009, pp. 325–347. ISSN: 15435938. DOI: 10.1146/ANNUREV-ENVIRON-011509-135201/CITE/REFWORKS. URL: https://www.annualreviews.org/content/journals/10.1146/annurev-environ-011509-135201.

Miguel, Edward and Michael Kremer (2004). "Worms: identifying impacts on education and health in the presence of treatment externalities". In: *Econometrica* 72.1, pp. 159–217. URL: https://doi.org/10.1111/j.1468-0262.2004.00481.x.

Miller, Grant and B. Piedad Urdinola (2010). "Cyclicality, Mortality, and the Value of Time: The Case of Coffee Price Fluctuations and Child Survival in Colombia". In: *Journal of Political Economy* 118.1, pp. 113–155.

Miller, Nora et al. (Nov. 2024). ""How I wish we could manage such things": A qualitative assessment of barriers to postpartum hemorrhage management and referral in Kenya". In: *PLOS Global Public Health* 4.11, pp. 1–16. DOI: 10.1371/journal.pgph.0003842. URL: https://doi.org/10.1371/journal.pgph.0003842.

Mireri, Junior (2020). *Stop Beating Kenyans During Curfew, Leaders Tell Police*. Accessed: 2025-04-12. The Standard. URL: https://www.standardmedia.co.ke/article/2001365989/stop-beating-kenyans-during-curfew-leaders-tell-police.

Mogstad, Magne, Andres Santos, and Alexander Torgovitsky (2018). "Using Instrumental Variables for Inference About Policy Relevant Treatment Parameters". In: *Econometrica* 86.5. ISSN: 1468-0262. DOI: 10.3982/ECTA15463. URL: https://onlinelibrary.wiley.com/doi/full/10.3982/ECTA15463%20https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA15463%20https://onlinelibrary.wiley.com/doi/10.3982/ECTA15463.

Moreira, Marcelo J. (July 2003). "A Conditional Likelihood Ratio Test for Structural Models". In: *Econometrica* 71.4, pp. 1027–1048. ISSN: 1468-0262. DOI: 10.1111/1468-0262.00438. URL: https://onlinelibrary.wiley.com/doi/full/10.1111/1468-0262.00438%20https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0262.00438%20https://onlinelibrary.wiley.com/doi/10.1111/1468-0262.00438.

Mwai, Daniel et al. (Mar. 2023). "Investment case for primary health care in low- and middle-income countries: A case study of Kenya". In: *PLOS ONE* 18.3, e0283156. ISSN: 1932-6203. DOI: 10.1371/JOURNAL.PONE.0283156. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0283156.

Nandi, Arijit et al. (2016). "Increased Duration of Paid Maternity Leave Lowers Infant Mortality in Low- and Middle-Income Countries: A Quasi-Experimental Study". In: *PLoS Medicine* 13.3, e1001985. DOI: 10.1371/journal.pmed.1001985. URL: https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001985.

Nareeba, T. et al. (2021). "Neonatal and child mortality data in retrospective population-based surveys compared with prospective demographic surveillance: EN-INDEPTH study". In: *Population Health Metrics* 19.Suppl 1, p. 7. DOI: 10.1186/s12963-020-00232-1.

Njuguna, John, Njoroge Kamau, and Charles Muruka (June 2017). "Impact of free delivery policy on utilization of maternal health services in county referral hospitals in Kenya". In: *BMC Health Services Research* 17 (1), pp. 1–6. ISSN: 14726963. DOI: 10.1186/S12913-017-2376-Z/TABLES/2. URL: https://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-017-2376-z.

Nkiaka, Ernest, Naveed R. Nawaz, and Jon C. Lovett (2022). "Review of Meteorological Drought in Africa: Historical Trends, Impacts, Mitigation Measures, and Prospects". In: *Pure and Applied Geophysics* 179.4, pp. 1365–1386. DOI: 10.1007/s00024-022-02988-z. URL: https://doi.org/10.1007/s00024-022-02988-z.

Novella, Rafael et al. (2012). *Conditional Cash Transfers, Female Bargaining Power and Parental Labour Supply.* Tech. rep. IDB-WP-368. Inter-American Development Bank.

Okeke, Edward N. (Mar. 2023). "When a Doctor Falls from the Sky: The Impact of Easing Doctor Supply Constraints on Mortality". en. In: *American Economic Review* 113.3, pp. 585–627. ISSN: 0002-8282. DOI: 10.1257/aer.20210701. URL: https://www.aeaweb.org/articles?id=10.1257/aer.20210701&from=f (visited on 03/07/2023).

Oo, Thae Aient Aient et al. (Dec. 2024). "Using the Health Belief Model to Understand Why Making Oral HIV Self-Testing Available to Truck Drivers in Kenya Had Little Impact on Six-Month Testing". In: *AIDS and Behavior* 28.12, pp. 4118–4126. ISSN: 15733254. DOI: 10.1007/S10461-024-04500-1/TABLES/3. URL: https://link.springer.com/article/10.1007/s10461-024-04500-1.

Orangi, Stacey et al. (2021). "Examining the implementation of the Linda Mama free maternity program in Kenya". In: *International Journal of Health Planning and Management* 36, pp. 2277–2296.

Ouellet, James V. and Vira Kasantikul (2006). "Motorcycle Helmet Effect on a Per-Crash Basis in Thailand and the United States". In: *https://doi.org/10.1080/15389580500338652* 7.1. ISSN: 10286586. DOI: 10.1080/15389580500338652. URL: https://www.tandfonline.com/doi/abs/10.1080/15389580500338652.

Ouko, Jacob Joseph Ochieng et al. (Aug. 2019). "Geographic information system-based evaluation of spatial accessibility to maternal health facilities in Siaya County, Kenya". In: *Geographical Research* 57.3, pp. 286–298.

Oyugi, Boniface et al. (Feb. 2023). "Does a free maternity policy in Kenya work? Impact and cost–benefit consideration based on demographic health survey data". In: *The European Journal of Health Economics.* ISSN: 1618-7598. DOI: 10.1007/s10198-023-01575-w.

Padmanabhan, V. and I.P.L. Png (Oct. 1995). "Returns Policies: Make Money by Making Good". In: *MIT Sloan Management Review.* URL: https://sloanreview.mit.edu/article/returns-policies-make-money-by-making-good/.

Pelnik, Carolyn (2024). "Moving to Profitability? Alleviating Constraints on Microentrepreneur Location". In: *Working Paper.*

Premji, S. (2014). "Perinatal distress in women in low- and middle-income countries: allostatic load as a framework to examine the effect of perinatal distress on preterm birth and infant health". In: *Maternal and Child Health Journal* 18.10. Erratum in: Matern Child Health J. 2015 Mar;19(3):691, pp. 2393–2407. DOI: 10.1007/s10995-014-1479-y.

Preston, Samuel H. (1975). "The changing relation between mortality and level of economic development". In: *Population Studies* 29.2, pp. 231–248.

Pryce, Joseph, Marty Richardson, and Christian Lengeler (Nov. 2018). "Insecticide-treated nets for preventing malaria". In: *The Cochrane Database of Systematic Reviews* 2018.11, p. CD000363. ISSN: 1469493X. DOI: 10.1002/14651858.CD000363.PUB3. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC6418392/.

Rao, M. R. et al. (2003). "Reliability of maternal recall and reporting of child births and deaths in rural Egypt". In: *Paediatric and Perinatal Epidemiology* 17.2, pp. 125–131. DOI: 10.1046/j.1365-3016.2003.00482.x.

Redfern, Alice et al. (2019). *Beneficiary preferences: Findings from Kenya and Ghana*. Tech. rep. IDinsight.

Republic of Kenya Ministry of Health (2025). *Facilities*. Accessed: 2025-06-04. URL: https://kmhfr.health.go.ke/public/facilities.

Riang'a, Rose M., Anne K. Nangulu, and Jacqueline E. W. Broerse (2018). "Perceived causes of adverse pregnancy outcomes and remedies adopted by Kalenjin women in rural Kenya". In: *BMC Pregnancy and Childbirth* 18.1, p. 408. DOI: 10.1186/s12884-018-2041-5. URL: https://bmcpregnancychildbirth.biomedcentral.com/articles/10.1186/s12884-018-2041-5.

Richterman, Aaron et al. (June 2023). "The effects of cash transfers on adult and child mortality in low- and middle-income countries". en. In: *Nature* 618.7965. Number: 7965 Publisher: Nature Publishing Group, pp. 575–582. ISSN: 1476-4687. DOI: 10.1038/s41586-023-06116-2. URL: https://www.nature.com/articles/s41586-023-06116-2 (visited on 11/28/2023).

Romano, Joseph P. and Michael Wolf (July 2005). "Stepwise Multiple Testing as Formalized Data Snooping". In: *Econometrica* 73 (4), pp. 1237–1282. ISSN: 1468-0262. DOI: 10.1111/J.1468-0262.2005.00615.X.

Romero Prieto, J., A. Verhulst, and M. Guillot (2021). "Estimating the infant mortality rate from DHS birth histories in the presence of age heaping". In: *PLoS One* 16.11, e0259304. DOI: 10.1371/journal.pone.0259304.

Rossin, Maya (2011). "The Effects of Maternity Leave on Children's Birth and Infant Health Outcomes in the United States". In: *Journal of Health Economics* 30.2, pp. 221–239. DOI: 10.1016/j.jhealeco.2011.01.005.

Rothschild, Michael (Oct. 1974). "A two-armed bandit theory of market pricing". In: *Journal of Economic Theory* 9.2, pp. 185–202. ISSN: 0022-0531. DOI: 10.1016/0022-0531(74)90066-0.

Rothschild, Michael and Joseph E. Stiglitz (Sept. 1970). "Increasing risk: I. A definition". In: *Journal of Economic Theory* 2.3, pp. 225–243. ISSN: 0022-0531. DOI: 10.1016/0022-0531(70)90038-4.

Ruhm, Christopher J. (2000). “Parental Leave and Child Health”. In: *Journal of Health Economics* 19.6, pp. 931–960. DOI: 10.1016/S0167-6296(00)00047-3.

Scorgie, F. et al. (2023). ““Mothers get really exhausted!” The lived experience of pregnancy in extreme heat: Qualitative findings from Kilifi, Kenya”. In: *Social Science & Medicine* 335, p. 116223. ISSN: 0277-9536. DOI: https://doi.org/10.1016/j.socscimed.2023.116223. URL: https://www.sciencedirect.com/science/article/pii/S0277953623005804.

Serina, Peter et al. (Dec. 2015). “Improving performance of the Tariff Method for assigning causes of death to verbal autopsies”. In: *BMC Medicine* 13.1, p. 291. ISSN: 1741-7015. DOI: 10.1186/s12916-015-0527-9. URL: https://doi.org/10.1186/s12916-015-0527-9 (visited on 11/28/2023).

Shrestha, Maheshwor (2020). “Get Rich or Die Tryin’: Perceived Earnings, Perceived Mortality Rates, and Migration Decisions of Potential Work Migrants from Nepal”. In: *The World Bank Economic Review* 34.1. ISSN: 0258-6770. DOI: 10.1093/WBER/LHZ023. URL: https://academic-oup-com.libproxy.berkeley.edu/wber/article/34/1/1/5588255.

Stedman, Nancy (2023). *Cash Transfer Programs Are Growing More Common in the U.S. as Studies Show They Improve People’s Health.* https://ldi.upenn.edu/our-work/research-updates/cash-transfer-programs-are-growing-more-common-in-the-u-s-as-studies-show-\\they-improve-peoples-health/. Accessed: 2025-02-23.

Stenberg, K. et al. (2021). “Cost-Effectiveness of Interventions to Improve Maternal, Newborn and Child Health Outcomes: A WHO-CHOICE Analysis for Eastern Sub-Saharan Africa and South-East Asia”. In: *International Journal of Health Policy and Management* 10.11, pp. 706–723. DOI: 10.34172/ijhpm.2021.07.

Strobl, Renate (Oct. 2022). “Background risk, insurance and investment behaviour: Experimental evidence from Kenya”. In: *Journal of Economic Behavior & Organization* 202, pp. 34–68. ISSN: 0167-2681. DOI: 10.1016/J.JEBO.2022.07.010.

Tanaka, Sakiko (2005). “Parental Leave and Child Health Across OECD Countries”. In: *The Economic Journal* 115.501, F7–F28. DOI: 10.1111/j.0013-0133.2005.00970.x.

United Nations (2015). *Goal 3: Ensure Healthy Lives and Promote Well-Being for All at All Ages.* Accessed: 2025-02-23. URL: https://www.un.org/sustainabledevelopment/health/.

— (2024). *World Population Prospects 2024: Standard Projections.*

United Nations Inter-agency Group for Child Mortality Estimation (2025). *All-Cause Child Mortality Data [dataset].* https://childmortality.org/all-cause-mortality/data. Accessed 2025-07-27.

United Nations Office for Disaster Risk Reduction (2023). *Horn of Africa Floods and Drought 2020–2023: Forensic Analysis.* Tech. rep. UNDRR. URL: https://www.undrr.org/resource/horn-africa-floods-and-drought-2020-2023-forensic-analysis.

Verhoogen, Eric (2021). “Firm-Level Upgrading in Developing Countries”. In: URL: http://www.nber.org/papers/w29461.

Walker, Michael W. et al. (Oct. 2024). "Slack and Economic Development". In: *NBER Working Paper*. DOI: 10.3386/W33055. URL: https://www.nber.org/papers/w33055.

Wiswall, Matthew and Basit Zafar (Feb. 2018). "Preference for the Workplace, Investment in Human Capital, and Gender". In: *The Quarterly Journal of Economics* 133.1, pp. 457–507. ISSN: 0033-5533. DOI: 10.1093/QJE/QJX035. URL: https://dx.doi.org/10.1093/qje/qjx035.

World Bank (2022). *Monitoring COVID-19 Impact on Households and Firms in Kenya*. Accessed: 2025-02-23. URL: https://www.worldbank.org/en/country/kenya/brief/monitoring-covid-19-impact-on-households-and-firms-in-kenya.

— (2024). *GDP per capita, PPP (constant 2021 international $)*. https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.KD.

World Health Organization (2022). *2022 WHO Verbal Autopsy Instrument*. Tech. rep. World Health Organization. URL: https://www.who.int/publications/m/item/2022-who-verbal-autopsy-instrument.

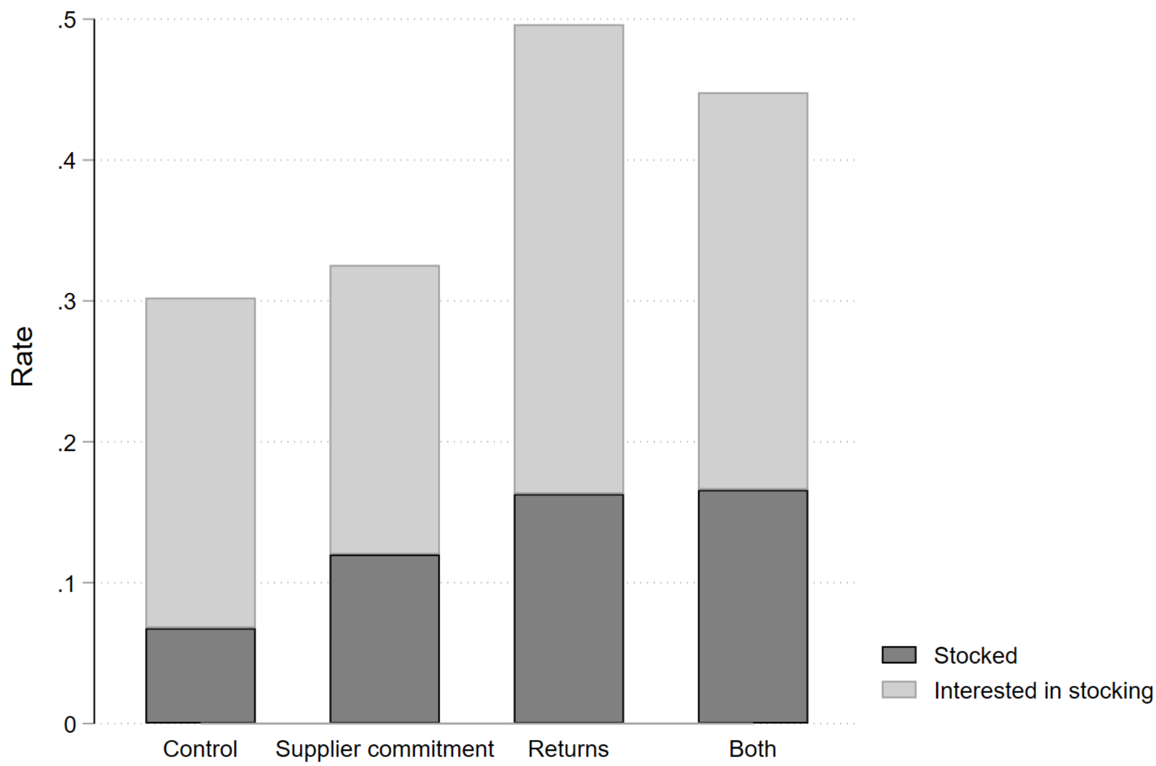
— (2025). *WHO Data — Kenya*. https://data.who.int/countries/404. WHO country code 404 corresponds to Kenya; accessed 2025-07-27.

Yaman, Fikriye Karanfil et al. (Aug. 2025). "The Impact of Advanced Maternal Age on Pregnancy Complications and Neonatal Outcomes". In: *Journal of clinical medicine* 14 (15). ISSN: 2077-0383. DOI: 10.3390/JCM14155387. URL: https://pubmed.ncbi.nlm.nih.gov/40807009/.

Appendix A

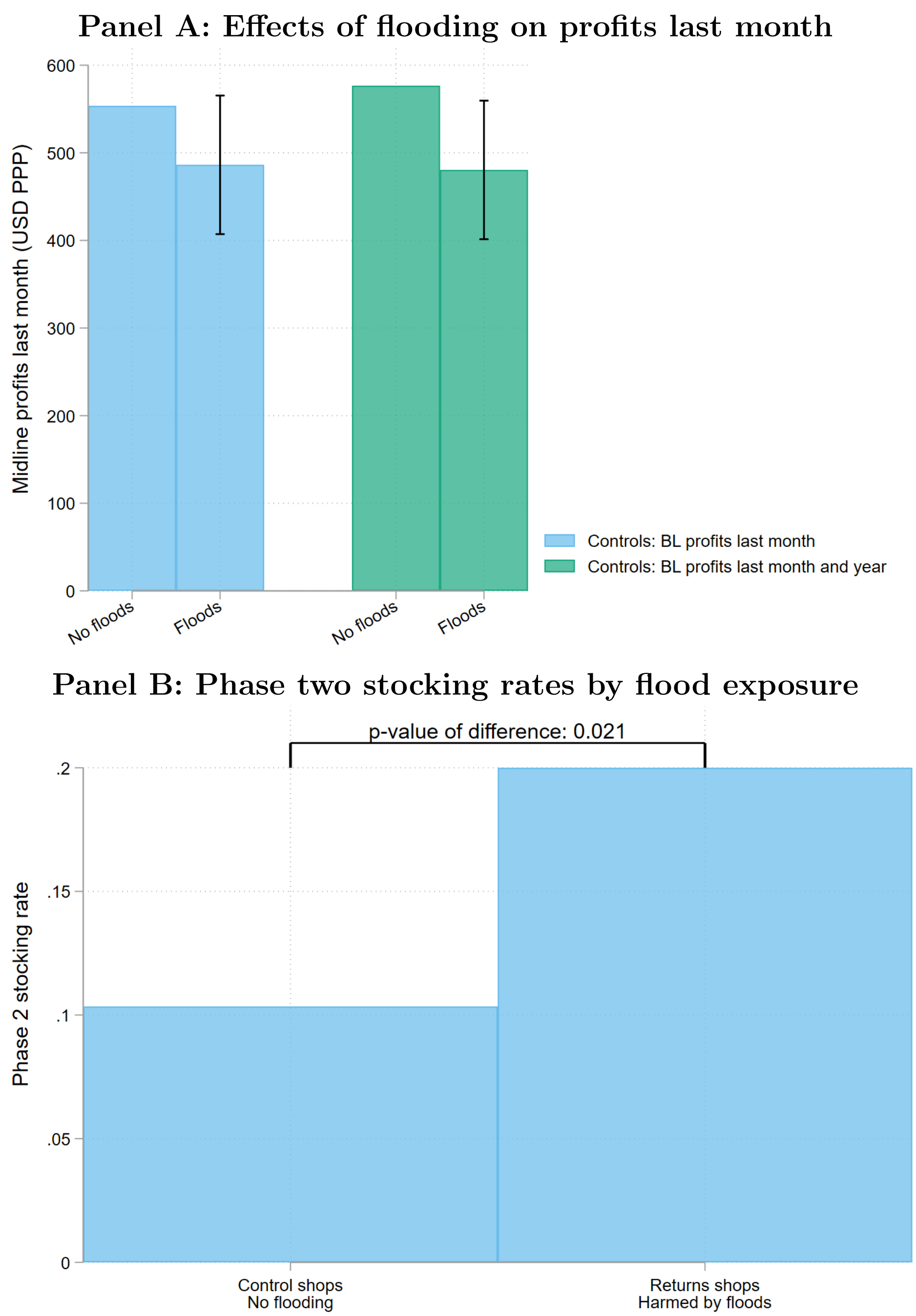
Appendix to Chapter 1

Figure A.1: Learning experiment: Treatment effects on phase one stocking and interest



This figure reports the rate of shops that stocked helmets in phase one, from the study or a different source, or that requested the field officer call them back in two days for a final purchase decision. The first bar presents the control stocking rate. The second reports the adoption rate among shops receiving the *supplier commitment*. The third bar is similar, but reports the effect of *returns*. The fourth bar reports the estimated effect of receiving both treatments.

Figure A.2: Learning experiment: Phase two stocking effects by flood exposure



Panel A plots profits the month before the midline among shops that reported they were harmed by flooding and those that were not. Results control for strata and industry fixed effects, controls for distance to the manufacturer, log revenue, days open per week, and indicators for stocking a new product in the year before the baseline survey, selling multiple products at baseline, and having space to store helmets without stocking less of another item. The first estimate also controls for baseline monthly profits, and the second controls for baseline monthly and annual profits. Panel B reports control, no flooding stocking and returns, flooding stocking rates. The p-value for equality is from a regression of stocking on returns among such firms, including the controls from profit estimates.

Table A.1: Summary statistics and baseline balance

	Insurance experiment		Learning experiment Returns		Learning experiment Supplier commitment	
Variable	(1) Control mean [SD]	(2) T - C	(3) Control mean [SD]	(4) T - C	(5) Control mean [SD]	(6) T - C
Respondent age	34.18	0.77	28.92	-0.77	28.68	-1.32
	[10.86]	(1.12)	[6.75]	(1.85)	[6.97]	(2.08)
Female	0.30	0.03	0.28	-0.03	0.27	0.00
	[0.46]	(0.05)	[0.45]	(0.03)	[0.44]	(0.03)
Years of education	12.78	0.19	13.23	0.16	13.34	-0.02
	[2.79]	(0.27)	[2.48]	(0.16)	[2.54]	(0.17)
Business age	4.53	-0.22	5.29	0.18	5.68	-0.57
	[5.46]	(0.52)	[5.45]	(0.36)	[5.77]	(0.36)
Motorcycle spares shop	0.59	-0.01	0.37	-0.02	0.37	-0.00
	[0.49]	(0.05)	[0.48]	(0.03)	[0.48]	(0.03)
Revenue last month	1,235.46	-132.01	1,459.28	153.17	1,590.50	-125.91
	[1,482.82]	(152.00)	[2,107.67]	(159.08)	[2,458.95]	(168.67)
Costs last month	972.27	15.70	655.92	138.05*	747.21	-49.55
	[1,280.78]	(185.11)	[835.61]	(79.85)	[1,045.04]	(83.74)
Profits last month	508.77	-46.39	622.90	61.17	674.88	-32.39
	[593.92]	(57.29)	[977.04]	(66.44)	[937.10]	(66.60)
1(paid employees)	0.22	0.05	0.25	-0.01	0.25	-0.01
	[0.41]	(0.04)	[0.43]	(0.03)	[0.43]	(0.03)
Wage bill last week	58.26	10.85	34.69	-5.05	21.60	18.33**
	[348.89]	(29.99)	[164.33]	(10.47)	[47.16]	(9.01)
Hours open/week	77.41	0.24	85.35	-0.16	86.06	-1.74*
	[19.50]	(1.57)	[15.37]	(0.95)	[15.07]	(0.95)
KM to helmet seller	NA	NA	2.24	0.06	2.28	-0.06
	–	–	[3.17]	(0.15)	[3.27]	(0.15)
Know helmet seller	0.56	-0.10**	0.25	0.03	0.27	0.00
	[0.50]	(0.05)	[0.44]	(0.03)	[0.44]	(0.03)
Min. to motorcycle taxi stand	NA	NA	3.16	-0.07	3.00	0.24
	–	–	[6.01]	(0.36)	[5.77]	(0.36)
New product in last year	0.28	0.03	0.34	-0.01	0.36	-0.05
	[0.45]	(0.04)	[0.48]	(0.03)	[0.48]	(0.03)
E[sales]	3.43	-0.04	3.75	0.06	3.89	-0.21
	[1.73]	(0.13)	[2.01]	(0.13)	[2.06]	(0.14)
V(sales)	2.96	-0.13	1.26	0.08	1.33	-0.06
	[2.88]	(0.25)	[1.84]	(0.12)	[1.80]	(0.12)
Observations	172	173	461	468	463	466
Joint p-value		0.87	0.91		0.47	

$* \ p < 0.1$, $** \ p < .05$ $*** \ p < .01$. Standard deviations in brackets. Standard errors in parenthesis.
Column (1) reports control means and standard deviations in the Insurance Experiment. Cols. (3) and (5) report control means and SDs in the Learning Experiment. Cols. (2), (4) and (6) report treatment - control means, estimated via OLS with strata fixed effects. P-values of joint orthogonality are estimated via seemingly unrelated regression.

Table A.2: Summary statistics and balance
Spillover survey

	Motorcycle shops		All shops	
	(1)	(2)	(3)	(4)
Variable	Control mean [SD]	T - C	Control mean [SD]	T - C
Respondent age	35.04	0.90	35.20	-0.00
	[10.41]	(0.99)	[10.59]	(0.82)
Business age	4.58	-0.33	4.95	-0.45
	[4.11]	(0.48)	[5.04]	(0.39)
Revenue last month	990.57	14.04	982.30	-60.27
	[759.23]	(101.57)	[805.51]	(82.68)
Costs last month	729.83	-53.75	710.49	-24.51
	[825.97]	(91.45)	[812.68]	(81.38)
Profits last month	465.55	-8.88	458.62	-27.24
	[342.76]	(50.30)	[349.98]	(41.93)
1(Employees)	0.30	-0.04	0.29	-0.03
	[0.46]	(0.06)	[0.45]	(0.04)
Helmet storage capacity	22.63	-10.77*	21.46	-2.12
	[57.87]	(6.07)	[51.71]	(6.64)
Owner hours of work/week	76.55	-1.30	76.71	-1.64
	[15.39]	(1.98)	[15.68]	(1.83)
Motorcycle spares shop	NA	NA	0.67	-0.20***
	–	–	[0.47]	(0.05)
Observations	219	157	327	338
Joint p-value		0.51		0.01

$* \ p < 0.1$, $** \ p < .05$ $*** \ p < .01$. Standard deviations in brackets. Standard errors in parenthesis clustered by market.

Column (1) reports the mean and standard deviation of the indicated variable across enterprises in pure control markets, restricting the sample to motorcycle spare part/repair shops. Column (3) is similar but with no sample restriction. Columns (2) and (4) report the difference between the treatment and control group in the sample and arm as denoted in the table header, estimated via OLS including county fixed effects. The last row reports the p-value associated with a test for joint orthogonality, constructed by estimating a seemingly unrelated regression model then estimating a Wald test to allow for missing variables. 35 shops that sold motorcycle helmets prior to the beginning of the study are excluded.

Table A.3: Baseline beliefs about helmet profitability

	Mean	SD	25th pctile	50th pctile	75th pctile	Obs
Panel A: Insurance experiment						
Pr(Loss| Stock 10 helmets)	0.307	0.233	0.100	0.300	0.500	350
Pr(Helmets profits > current stock)	0.504	0.219	0.400	0.500	0.600	349
Pr(Sell out in 8 weeks|Stock)	0.442	0.256	0.200	0.450	0.600	350
$\mathbb{E}$[8 week revenue] − stock cost	0.200	16.189	-8.850	1.996	10.672	340
8 week expected sales	3.436	1.800	2.050	3.250	4.575	340
8 week SD sales	1.489	0.824	0.829	1.303	2.041	340
Panel B: Learning experiment						
Pr(Loss | Buy 10 helmets)	0.369	0.241	0.200	0.300	0.500	922
Pr(Helmets profits > current stock)	0.506	0.214	0.400	0.500	0.600	922
Pr(Representative firm restocked)	0.442	0.205	0.333	0.444	0.556	929
$\mathbb{E}$[1 month revenue] − stock cost	-5.078	29.082	-28.330	-3.384	16.139	873
1 month expected sales	3.765	2.052	2.300	3.500	5.000	873
1 month SD sales	0.928	0.658	0.500	0.829	1.221	873

This table reports baseline beliefs about helmet profitability. The first row in each panel is the firm's belief about the likelihood that they would lose money, inclusive of the opportunity cost of funds, if they stocked 10 helmets. The second row is the likelihood that stocking 10 helmets would raise the firm's profits, inclusive of any losses from stocking less of other items. The third row in Panel A denotes the firm's belief about their probability of selling 3 helmets in 8 weeks if they stocked them, and in Panel B the row denotes the enterprise's perceived likelihood that a representative shop would restock if given helmets to learn about the market. The fourth row captures expected revenue net of stocking cost for the study offers over an 8 week period in Panel A and a 1 month period in Panel B. The final two rows present beliefs about expected sales and the standard deviation of sales, measured with a frequentist mechanism, over an 8 week and 1 month period respectively.

Table A.4: Learning experiment: Heterogeneous effects by presence of a pre-existing seller

	(1)	(2)	(3)	(4)
	Stocked in Phase 1		Stocked in Phase 2	
Returns	0.127***	0.126***	0.078**	0.077**
	(0.032)	(0.031)	(0.034)	(0.033)
Supplier commitment	0.085***	0.085***	0.069**	0.055*
	(0.029)	(0.028)	(0.032)	(0.030)
Returns × Supplier commitment	-0.099**	-0.079*	-0.080	-0.051
	(0.048)	(0.046)	(0.049)	(0.047)
BL seller	0.125**	0.092*	0.092*	0.029
	(0.050)	(0.047)	(0.053)	(0.048)
Returns × BL seller	-0.122	-0.143**	-0.038	-0.047
	(0.075)	(0.072)	(0.080)	(0.074)
Supplier commitment × BL seller	-0.128*	-0.099	-0.112	-0.043
	(0.073)	(0.071)	(0.077)	(0.071)
Returns × Commitment	0.166	0.142	0.074	0.019
× BL seller	(0.107)	(0.104)	(0.111)	(0.106)
Observations	929	929	929	929
Control mean	0.034	0.034	0.073	0.073
Controls	Strata FE	Yes	Strata	Yes

$* \ p < 0.1, \ \ ** \ p < .05, \ \ ***p < .01$

This table plots heterogeneous treatment effects on helmet stocking and helmet market entry from the Learning Experiment based on proximity to an existing helmet seller. "BL Seller" indicates that a shop reported that a pre-existing helmet seller operated near them at baseline. Columns (1)-(2) reports effects on phase 1 stocking, supplied by the study or any other source. Column (3)-(4) report phase 2 stocking effects. The dependent variable in columns (3)-(4) is coded to 0 if a shop withdrew from the study because they did not wish to sell helmets or if the outcome was confirmed without the survey. All estimates include strata fixed effects. Columns (2) and (4) include industry fixed effects, controls for distance to the manufacturer, log revenue, days open per week, and indicators for stocking a new product in the year before the baseline survey, selling multiple products at baseline, and having space to store helmets without stocking less of another item.

Table A.5: Learning experiment: Belief-adoption relationship, robustness

	Sample: Pr(Loss|Stock 5) > 0			All	Non-deg.	All
	(1) Revenue	(2) Sales	(3) Sales	(4) Sales	(5) Sales	(6) Sales
Returns	-0.193	-0.006	0.024	0.015	0.022	-0.022
	(0.213)	(0.073)	(0.072)	(0.067)	(0.086)	(0.076)
$\mathbb{E}$[sales/revenue]	0.052*	0.021	0.021	0.008	0.023*	0.023
	(0.028)	(0.015)	(0.015)	(0.008)	(0.013)	(0.015)
σ(sales/revenue)	-0.041**	-0.042**	-0.031*	-0.022	-0.054**	-0.054**
	(0.020)	(0.020)	(0.018)	(0.017)	(0.024)	(0.022)
Returns × $\mathbb{E}$[sales/revenue]	-0.001	0.014	0.008	0.013	-0.018	0.013
	(0.065)	(0.033)	(0.033)	(0.015)	(0.021)	(0.033)
Returns × σ(sales/revenue)	0.045	0.039	0.032	0.023	0.092*	0.051
	(0.038)	(0.053)	(0.053)	(0.039)	(0.053)	(0.054)
Pr(Loss|Stock 5) = 0						-0.071
						(0.109)
Returns × Pr(Loss|Stock 5) = 0						0.239
						(0.220)
$\mathbb{E}$[sales/revenue] × Pr(Loss|Stock 5) = 0						-0.010
						(0.024)
σ(sales/revenue) × Pr(Loss|Stock 5) = 0						0.101
						(0.086)
Returns × $\mathbb{E}$[sales/revenue] × Pr(Loss|Stock 5) = 0						-0.022
						(0.047)
Returns × σ(sales/revenue) × Pr(Loss|Stock 5) = 0						-0.136
						(0.146)
Observations	283	304	304	427	286	427
Controls	Full	DPLASSO	None	Full	Full	Full

$*\ p < 0.1$, $**\ p < .05$ $***\ p < .01$. Robust standard errors in parenthesis. The dependent variable is 1 if the firm stocked helmets in phase 1. *Supplier commitment* firms are excluded. In column (1), the independent variables are log(1 + $\mathbb{E}$[helmet revenue]) and log(1 + $\mathbb{V}$[helmet revenue]). Columns (2)-(6) examine $\mathbb{E}[sales]$ and SD[sales]. Columns (1)-(3) exclude shops reporting 0 loss risk from stocking 5 helmets. Columns (4)-(6) include all observations, but 5 screens for comprehension of belief elicitation by dropping respondents reporting degenerate beliefs (Var(sales) ≤ 0.25). Covariates in column (2) were selected with double post LASSO. Columns (1) and (4)-(6) include controls for storage space, education, baseline profits, respondent characteristics, and firm characteristics.

Table A.6: Learning experiment: Beliefs of adopters with versus without returns

	(1) Belief levels	(2) Belief logs
$\mathbb{E}$[sales/revenue]	-0.001	
	(0.038)	
σ(sales/revenue)	0.202*	
	(0.111)	
log(1 + E[sales])		-0.012
		(0.175)
log(1 + V(sales))		0.246**
		(0.113)
Know helmet seller	-0.309**	-0.313**
	(0.148)	(0.140)
Observations	46	46

$* \ p < 0.1, \quad ** \ p < .05 \quad *** \ p < .01$

Bootstrapped standard errors in parenthesis (500 draws). This table reports beliefs about demand for helmets of adopters with and without *returns*. The dependent variable in each column is an indicator equal to 1 if the agent received *returns*, and the sample is restricted to those that stocked helmets in phase one and did not receive the supplier commitment offer. Robust standard errors in parenthesis.

Table A.7: Learning experiment: Helmets sales outcomes vs expectations (phase 1 adopters)

Panel A: Levels				
	Sales		Revenue	
	(1)	(2)	(3)	(4)
	BL → ML	ML → EL	BL → ML	ML → EL
Expectation at BL	0.169		0.199	
	(0.123)		(0.159)	
Expectation at ML		0.544***		0.463**
		(0.119)		(0.217)
Constant	3.229***	1.199***	67.270***	29.601**
	(0.603)	(0.431)	(16.052)	(14.219)
Observations	107	111	109	116
Adjusted R^2	0.011	0.179	0.007	0.073
Pr(η BL= η ML)		0.032		0.300
Panel B: Elasticities				
	log(Sales)		log(Revenue)	
	(1)	(2)	(3)	(4)
	BL → ML	ML → EL	BL → ML	ML → EL
log(Expectation at BL)	0.279**		0.239*	
	(0.122)		(0.131)	
log(Expectation at ML)		0.690***		0.364***
		(0.139)		(0.114)
Constant	0.921***	0.425***	3.298***	2.703***
	(0.184)	(0.139)	(0.587)	(0.463)
Observations	98	91	97	85
Adjusted R^2	0.045	0.239	0.028	0.099
Pr(η BL= η ML)		0.047		0.473

$* \ p < 0.1$, $** \ p < .05$ $*** \ p < .01$. Robust standard errors in parenthesis. This table reports regressions of sales outcomes and revenue on expectations elicited from respondents before the indicated period. Panel A reports estimates in levels and Panel B in elasticities. Includes only observations where the shop stocked helmets in phase 1. Columns (1)-(2) report beliefs about sales while columns (3)-(4) consider beliefs about revenue, constructed from expected sales and expected prices. Beliefs were elicited using a frequentist method. Columns (1) and (3) report regressions of the log of sales or revenue from baseline to midline on log baseline expectations of this value. Columns (2) and (4) report regressions of the log of sales or revenue from midline to endline on log midline expectations of this value. The p-values in columns (2) and (4) report results of a test that the baseline and midline elasticities are equal, estimated using seemingly unrelated regressions.

Table A.8: Learning experiment: Helmet profit dynamics and selling costs

	(1) Helmet profits per month EL	(2) Total costs selling helmets	(3) Helmet costs net est. stock	(4) Any fixed costs	(5) Fixed costs	(6) Δ capacity utilization
Stocked in phase 1	-9.419					0.046*
	(5.956)					(0.024)
Estimated cost of helmet stock (PPP)		0.451***				
		(0.167)				
Returns			-8.747	0.050	-0.632	
			(8.575)	(0.060)	(2.009)	
Observations	127	132	132	128	128	712
Dep. var. mean	21.552	73.735	-10.925	0.133	2.381	0.064

$* \ p < 0.1, \ ** \ p < .05 \ *** \ p < .01$

Robust standard errors in parenthesis. The sample consists of shops that reported selling helmets, supplied by the study or a different source. Column (1) regresses helmet profits accumulated between midline and endline on an indicator equal to 1 if the shop stocked helmets in the month after baseline to test for learning by doing. Column (2) regresses total reported costs of stocking helmets on the cost of the helmet stock, measured via administrative data or estimated. The dependent variable in column (3) equals reported helmet costs net of stock price. Column (4) examines whether shops reported any fixed costs of helmet sales, and column (5) reports total fixed costs. Column (6) reports the change in capacity utilization of the firm from baseline to endline, measured as worst week over best week profits in the last month.

Table A.9: Learning experiment: Helmet adoption rates and proximity of other sellers

	(1) Ordered Baseline	(2) Ordered Midline	(3) Ordered Endline	(4) Ordered	(5) Observed Seller
Know helmet seller	0.129**				0.325***
	(0.055)				(0.034)
Know of large seller, BL	-0.198***				
	(0.068)				
Noticed seller by ML		0.179**			
		(0.085)			
Noticed helmet seller by EL			0.201***		
			(0.063)		
Study adopter within 0.25km				0.031**	0.138***
				(0.014)	(0.031)
Sample shops in quarter km				-0.001***	
				(0.000)	
Study adopter 0.25-0.5km					-0.003
					(0.030)
Observations	231	165	165	771	929
Control mean	0.034	0.028	0.029	0.010	0.237
Controls	Lasso	Lasso	Lasso	Lasso	No
Sample restriction	Control	No BL order control	No BL order control	No BL order	

$* \; p < 0.1$, $** \; p < .05$, $*** p < .01$

Robust standard errors in parenthesis. This table reports correlations between proximity to a shop selling helmets and the firm's own stocking decisions. In column (1), the dependent variable equals 1 if the shop stocked in phase one, in column (2) the dependent variable captures stocking between a month after baseline and a month after midline, column (3) looks at stocking more than a month after midline, and column (4) captures ever stocking. Column (5) indicates that the shop reported knowing a helmet seller in their location at midline. Noticed seller indicates that the shop did not know of a shop selling helmets before the indicated survey, then observed one. Columns (1) - (3) include only untreated shops, with (2) and (3) further excluding those that stocked in the 4 weeks after baseline. Column (4) includes all firms that did not stock helmets within a month of baseline. Column (1)-(4) includes enterprise and shopkeeper demographic controls selected using double-post selection LASSO (Belloni, Chernozhukov, and Hansen, 2014a).

Table A.10: Learning experiment: Effect of returns on helmet access

	(1)	(2)	(3)	(4)	(5)	(6)
	Helmet shop in market, endline	Ever seller in market	Cum. helmets stocked, endline		Cum. helmets sold, endline	
Returns	0.077**	0.135***	1.870**	2.828***	1.374*	2.111**
	(0.038)	(0.040)	(0.835)	(1.014)	(0.747)	(0.951)
Supplier commitment	0.044	0.083**	0.663	1.281***	0.322	0.763**
	(0.038)	(0.042)	(0.480)	(0.460)	(0.348)	(0.343)
Returns × Commitment	-0.035	-0.096	-1.455	-2.745**	-1.238	-2.105**
	(0.054)	(0.058)	(0.978)	(1.159)	(0.829)	(1.043)
BL Seller				2.013		1.492
				(1.322)		(0.967)
Returns × BL Seller				-3.651**		-2.828**
				(1.750)		(1.357)
Commitment × BL Seller				-2.508		-1.817
				(1.579)		(1.141)
Returns × Commitment × BL seller				4.876**		3.327**
				(2.359)		(1.690)
Observations	929	929	929	929	926	926
Control mean	0.242	0.314	1.525	0.689	0.966	0.407

$* \ p < 0.1$, $** \ p < .05$, $***p < .01$

Robust standard errors in parenthesis. This table reports evidence of the effect of the intervention on helmet access. Column (1) equals 1 if the respondent shop entered the helmet market by endline or reported that a shop near them sold helmets. Column (2) is similar but equals 1 if the respondent ever stocked helmets or ever reported that a shop near them stocked helmets. Columns (3) - (4) examine the total number of helmets stocked by the shop by endline, and columns (5) - (6) examine total helmet sales by endline. In 3 cases, shops stocked helmets and did not complete the endline survey because their enterprises closed, so endline sales are imputed using midline values. Estimates include strata fixed effects and controls for industry, proximity to the manufacturer, days open per week at baseline, log baseline revenue, indicators for adopting a new product in the year before the survey, selling multiple products at baseline and having space to store helmets. I also control for exposure to floods at midline.

A.1 Model Details

Lagrangian and solution to optimal investment

The Lagrangian of the entrepreneur's optimization problem may be written as

$$\begin{aligned}\mathcal{L} = {} & u(c_1) + \lambda_1\left[(1+r)a_0 - c_1 - a_1 - w_s I_{s1} - w_n I_{n1}\right] \\ & + \kappa_{a1}\left[a_1 - \bar{a}\right] + \kappa_{\chi 1}\left[I_{n1}\cdot(I_{n1}-\chi)\right] + \iota_{s1} I_{s1} \\ & + \delta\,\mathbb{E}_{\theta,\nu_{s2},\nu_{n2}}\left\{u(c_2) + \lambda_2\left[\pi_s(I_{s1},\nu_{s2}) + \pi_n(I_{n1},\nu_{n2}+\theta) + (1+r)a_1 - c_2 - a_2 - w_s I_{s2} - w_n I_{n2}\right]\right. \\ & \left. + \kappa_{a2}\left[a_2 - \bar{a}\right] + \kappa_{\chi 2}\left[I_{n2}\cdot(I_{n2}-\chi)\right] + \iota_{s2} I_{s2} | \mathcal{I}_1\right\} \\ & + \delta^2\,\mathbb{E}_{\nu_{s2},\nu_{n2},\theta}\left[V^*(y_2,\theta) - \bar{R}(y_2,\mathcal{I}_2(I_{n1})|\mathcal{I}_1\right]\end{aligned} \tag{A.1}$$

Differentiating, we get the set of first order conditions

$$\begin{aligned} & \mathcal{L}_c : u'(c_t) = \lambda_t \\ & \mathcal{L}_a : \lambda_t = \mathbb{E}_t\,\lambda_{t+1} + \kappa_{at} \\ & \mathcal{L}_{I_s} : \lambda_t w_s + \iota_{st} = \delta\,\mathbb{E}_t\left[\lambda_{t+1}\frac{\partial}{\partial I_{st}}\pi_s(I_{st},\nu_{st+1})\right] \\ & \mathcal{L}_{I_n} : \lambda_t w_n + \kappa_{\chi t}(2I_{nt}-\chi) = \delta\,\mathbb{E}_t\left[\lambda_{t+1}\frac{\partial}{\partial I_{nt}}\pi_n(I_{nt},\nu_{nt+1}+\theta)\right] - \delta^2\,\mathbb{E}_t\left[\frac{\partial}{\partial I_{nt}}\bar{R}(y_{t+1},\mathcal{I}_{t+1}(I_{nt}))\right]\end{aligned} \tag{A.2}$$

where $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot|\mathcal{I}_t]$.

From the FOCs for consumption and assets, we get the consumption Euler equation $u'(c_t) = \mathbb{E}_t\,u'(c_{t+1}) + \kappa_{at}$. By Karush–Kuhn–Tucker conditions, whenever capital constraints are not binding, $u'(c_t) = \mathbb{E}_t\,u'(c_{t+1})$, and whenever they bind $\kappa_{at} > 0 \Rightarrow u'(c_t) > \mathbb{E}_t\,u'(c_{t+1})$, so $u'(c_t) \geq \mathbb{E}_t\,u'(c_{t+1})$.

Next solving for optimal investment in the safe good,

$$\begin{aligned}\delta\,\mathbb{E}_t\left[\lambda_{t+1}\frac{\partial}{\partial I_{st}}\pi_s(I_{st},\nu_{st+1})\right] &= \lambda_t w_s + \iota_{st} \\ \delta\,\mathbb{E}_t\left[\frac{u'(c_{t+1})}{u'(c_t)}\frac{\partial}{\partial I_{st}}\pi_s(I_{st},\nu_{st+1})\right] &= w_s + \frac{1}{u'(c_t)}\iota_{st} \\ \delta\left\{\mathbb{E}_t\left[\frac{u'(c_{t+1})}{u'(c_t)}\right]\mathbb{E}_t\left[\frac{\partial}{\partial I_{st}}\pi_s(I_{st},\nu_{st+1})\right] + \right. & \\ \left.\frac{1}{u'(c_t)}Cov_t\left(u'(c_{t+1}),\frac{\partial}{\partial I_{st}}\pi_s(I_{st},\nu_{st+1})\right)\right\} &= w_s + \frac{1}{u'(c_t)}\iota_{st}\end{aligned}$$

where ι_{st} is a Lagrangian multiplier ensuring non-negative investment, which will not bind and be zero whenever the safe product is stocked. Optimal investment in the new good is

$$\delta \left\{ \mathbb{E}_t \left[\frac{u'(c_{t+1})}{u'(c_t)} \right] \mathbb{E}_t \left[\frac{\partial}{\partial I_{nt}} \pi_n(I_{nt}, \nu_{nt+1} + \theta) \right] + \frac{1}{u'(c_t)} Cov_t \left(u'(c_{t+1}), \frac{\partial}{\partial I_{nt}} \pi_n(I_{nt}, \nu_{nt+1}) \right) - \frac{1}{u'(c_t)} \delta \, \mathbb{E}_t \left[\frac{\partial}{\partial I_{nt}} \bar{R}(y_{t+1}, \mathcal{I}_{t+1}(I_{nt})) \right] \right\} = w_n + \frac{1}{u'(c_t)} \iota_{\chi t}(2I_{nt} - \chi)$$

given that this value satisfies the participation constraint that the utility exceeds that from investing 0, where this verification is needed since minimum order sizes create a non-convex choice set. This Euler equation characterizes optimal investment given adoption. Any change in primitives that raises the optimal inventory level given entry, I^*_{nt}, strictly raises the value of entry, and thus weakly increases the probability of stocking. Therefore comparative statics on this Euler equation are sufficient to study entry probability.

Derivation of the derivative of Bayesian regret w.r.t investment

This section shows that

$$\frac{\partial}{\partial I_{nt}} \mathbb{E} \left[\bar{R}(y_{t+1}, \mathcal{I}_{t+1}(I_{nt})) | \mathcal{I}_t \right] = -\frac{1}{2} \sum_{\tau=t+2}^{\infty} \delta^{\tau-2} Cov \left(R_\tau(\theta), (\theta - \mu_\tau)' I_{nt}^{-2} V_{nt}^{-1} \Sigma_x V_{nt}^{-1} (\theta - \mu_\tau) \middle| \mathcal{I}_t \right) \leq 0$$

where $V_{nt} = I_{nt}^{-1}\Sigma_x + \Sigma_n$. The inequality is typically strict whenever expected regret is positive.

Consider an agent in time t making investment I_{nt}. In period $t+1$, they will receive the signal from this investment, which will affect their Bayesian regret associated with decisions beginning in period $t+2$, $\bar{R}(y_{t+1}, \mathcal{I}_{t+1}(I_{nt}))$. The aim is to find

$$\frac{\partial}{\partial I_{nt}} \mathbb{E} \left[\bar{R}(y_{t+1}, \mathcal{I}_{t+1}(I_{nt})) | \mathcal{I}_t \right]$$

Let $c^*_\tau(x_t, \theta)$ denote the expected utility maximizing consumption path given θ and $\bar{c}_\tau(x_t, \mathcal{I}_{nt+1}, \theta)$ be consumption along the agent's planned expected utility maximizing path if θ is the true parameter. Applying the definition of Bayesian regret and the Law of Iterated Expectations

$$\bar{R}(y_{t+1}, \mathcal{I}_{t+1}(I_{nt})) = \sum_{\tau=t+2}^{\infty} \delta^{\tau-t-1} \mathbb{E}_{\mathcal{I}_\tau, \nu_n, \nu_s} \left[\int_{\mathbb{R}^k} (u(c^*_\tau(\theta)) - u(\bar{c}_\tau(\theta))) f(\theta | \mathcal{I}_\tau(x(I_{nt}))) d\theta \middle| \mathcal{I}_{t+1}, x(I_{nt}) \right]$$

The expectation over $\mathcal{I}_\tau$ captures expected future learning about demand, due to planned investment along the consumption path or learning from neighbors. $x(I_{nt})$ gives the realized draw of x given I_{nt}. Plugging this in

$$\mathbb{E}\left[\bar{R}(y_{t+1}, \mathcal{I}_{t+1}(I_{nt}))|\mathcal{I}_t\right] =$$
$$\mathbb{E}_{\mathcal{I}_\tau,\nu_n,\nu_s}\left[\sum_{\tau=t+2}^{\infty} \delta^{\tau-t-1} \int_{\mathbb{R}^k}\int_{\mathbb{R}^k} (u(c^*_\tau(\theta)) - u(\bar{c}_\tau(\theta)))f(\theta|\mathcal{I}_\tau(x(I_{nt})))f(x(I_{nt})|\mathcal{I}_t)\ d\theta\ dx(I_{nt})\Big|\mathcal{I}_t\right]$$

The integral over $x(I_{nt})$ captures the fact that at time t, the agent does not know what draw $x(I_{nt})$ they will receive, so they must take an expectation over the possible signals. Focusing on some arbitrary τ,

$$\frac{\partial}{\partial I_{nt}}\bar{R}_\tau \equiv \frac{\partial}{\partial I_{nt}}\mathbb{E}_{\mathcal{I}_\tau,\nu_n,\nu_s}\left[\int_{\mathbb{R}^k}\int_{\mathbb{R}^k} (u(c^*_\tau(\theta)) - u(\bar{c}_\tau(\theta)))f(\theta|\mathcal{I}_\tau(x(I_{nt})))f(x(I_{nt})|\mathcal{I}_t)\ d\theta\ dx(I_{nt})\Big|\mathcal{I}_t\right]$$

where $f(\theta|\mathcal{I}_\tau(x(I_{nt})))$ is conditioning on the draw of $x(I_{nt})$. I apply the Law of Iterated Expectations and integrate over the distribution of expected draws $f(x(I_{nt})|\mathcal{I}_t)$. The notation $\mathbb{E}_{\mathcal{I}_\tau}$ refers to expected draws to the information set, capturing expected evolution aside from the signal $x(I_{nt})$, including information from peers and other planned investments in the new product. By the Envelope Theorem,

$$\begin{aligned}\frac{\partial}{\partial I_{nt}}\bar{R}_\tau = \mathbb{E}_{\mathcal{I}_\tau,\nu_n,\nu_s}\Bigg[&\int_{\mathbb{R}^k}\int_{\mathbb{R}^k} (u(c^*_\tau(\theta)) - u(\bar{c}_\tau(\theta)))\frac{\partial f(\theta|\mathcal{I}_\tau(x(I_{nt})))}{\partial I_{nt}}f(x(I_{nt})|\mathcal{I}_t)\ d\theta\ dx(I_{nt})\\ &+ \int_{\mathbb{R}^k}\int_{\mathbb{R}^k} (u(c^*_\tau(\theta)) - u(\bar{c}_\tau(\theta)))f(\theta|\mathcal{I}_\tau(x(I_{nt})))\frac{\partial f(x(I_{nt})|\mathcal{I}_t)}{\partial I_{nt}}\ d\theta\ dx(I_{nt})\Big|\mathcal{I}_t\Bigg]\end{aligned}$$

The agent knows $x(I_{nt}) \sim \mathcal{N}(\theta_0, V_{nt})$ where $V_{nt} = I_{nt}^{-1}\Sigma_x + \Sigma_n$, but does not know θ_0. Expected signal draws are thus distributed $\mathcal{N}(\mu_t, \Sigma_t + V_{nt})$ where μ_t and Σ_t are posteriors given $\mathcal{I}_t$. I first show that the second term is 0, which is intuitively the case since a change in investment level should not change the location of the signal. Formally,

$$\begin{aligned}\int_{\mathbb{R}^k}\frac{\partial f_\tau(x(I_{nt}))}{\partial I_{nt}}dx(I_{nt}) &= \int_{\mathbb{R}^k}\frac{1}{2}I_{nt}^{-2}f(x(I_{nt})|\mathcal{I}_t)\left\{Tr\left([V_{nt}+\Sigma_t]^{-1}\Sigma_n\right)\right.\\ &\quad \left.-\left[(x-\mu_t)'\left[V_{nt}+\Sigma_t\right]^{-1}\Sigma_n\left[V_{nt}+\Sigma_t\right]^{-1}(x-\mu_t)\right]\right\}dx(I_{nt})\\ &= \frac{1}{2}I_{nt}^{-2}\left\{Tr\left([V_{nt}+\Sigma_t]^{-1}\Sigma_n\right) - Tr\left([V_{nt}+\Sigma_t]^{-1}\Sigma_n\right)\right\} = 0\end{aligned}$$

where the first line applies Jacobi's formula for the derivative of a trace and the second line leverages results about the expectation of a quadratic form.

Now I turn to the first term in the derivative. This will be non-zero since the agent expects investment to lower posterior variance, which affects $u(c^*_\tau(\theta)) - u(\bar{c}_\tau(\theta))$. For a fixed draw $x(I_{nt})$,

$$\frac{\partial f(\theta|\mathcal{I}_\tau(x(I_{nt})))}{\partial I_{nt}} = -\frac{1}{2}Tr\left(\Sigma_\tau^{-1}\frac{\partial \Sigma_\tau}{\partial I_{nt}}\right) f(\theta|\mathcal{I}_\tau) + \frac{1}{2}\left[(\theta-\mu_\tau)'\Sigma_\tau^{-1}\frac{\partial \Sigma_\tau}{\partial I_{nt}}\Sigma_t^{-1}(\theta-\mu_\tau)\right] f(\theta|\mathcal{I}_\tau)$$
$$+ (\theta-\mu_\tau)'\Sigma_\tau^{-1}\frac{\partial \mu_\tau}{\partial I_{nt}} f(\theta|\mathcal{I}_\tau)$$

From the known form of the posterior mean and variance, it follows that

$$\frac{\partial \mu_\tau}{\partial I_{nt}} = I_{nt}^{-2}\Sigma_\tau V_{nt}^{-1}\Sigma_x V_{nt}^{-1}(x(I_{nt})-\mu_\tau)$$
$$\frac{\partial \Sigma_\tau}{\partial I_{nt}} = -I_{nt}^{-2}\Sigma_\tau V_{nt}^{-1}\Sigma_x V_{nt}^{-1}\Sigma_\tau$$

Terms with $\frac{\partial \mu_\tau}{\partial I_{nt}}$ will be zero, reflecting the fact that the agent doesn't expect a marginal increase in investment to change the location parameter of beliefs. In particular,

$$\mathbb{E}_\theta\left[(u(c_\tau^*(\theta)) - u(\bar{c}_\tau(\theta)))(\theta-\mu_\tau)'\Sigma_\tau^{-1}\frac{\partial \mu_\tau}{\partial I_{nt}}\big|x(I_{nt})\right]$$
$$= \mathbb{E}_\theta\left[(u(c_\tau^*(\theta)) - u(\bar{c}_\tau(\theta)))\big|x(I_{nt})\right]\underbrace{\mathbb{E}_\theta\left[(\theta-\mu_\tau)'\big|x(I_{nt})\right]}_{=0}\Sigma_\tau^{-1}\frac{\partial \mu_\tau}{\partial I_{nt}}$$
$$+ Cov\left[(u(c_\tau^*(\theta)) - u(\bar{c}_\tau(\theta))), (\theta-\mu_\tau)'\big|x(I_{nt})\right]\Sigma_\tau^{-1}\frac{\partial \mu_\tau}{\partial I_{nt}}$$
$$= Cov\left[(u(c_\tau^*(\theta)) - u(\bar{c}_\tau(\theta))), (\theta-\mu_\tau)'\big|x(I_{nt})\right] I_{nt}^{-2}V_{nt}^{-1}\Sigma_x V_{nt}^{-1}\Sigma_\tau(x(I_{nt})-\mu_\tau)$$

Since $\mathbb{E}_{\mathcal{I}_\tau,\nu_n,\nu_t}[x(I_{nt})-\mu_\tau]=0$, it follows that

$$\mathbb{E}_{\mathcal{I}_\tau,\nu_n,\nu_t}\left[\mathbb{E}_\theta\left[(u(c_\tau^*(\theta)) - u(\bar{c}_\tau(\theta)))(\theta-\mu_\tau)'\Sigma_\tau^{-1}\frac{\partial \mu_\tau}{\partial I_{nt}}\big|x(I_{nt})\right]\right] = 0$$

Now denote $P_{nt} = V_{nt}^{-1}$ so that $\frac{\partial P_{nt}}{\partial I_{nt}} = I_{nt}^{-2}P_{nt}\Sigma_x P_{nt}$. Substituting for $\frac{\partial \Sigma_\tau}{\partial I_{nt}}$, collecting the terms containing this expression, and evaluating

$$\mathbb{E}_\theta\left[(u(c_\tau^*(\theta)) - u(\bar{c}_\tau(\theta)))\cdot\left(\frac{1}{2}Tr\left(\frac{\partial P_{nt}}{\partial I_{nt}}\Sigma_\tau\right) - \frac{1}{2}(\theta-\mu_\tau)'\frac{\partial P_{nt}}{\partial I_{nt}}(\theta-\mu_\tau)\right)\big|x(I_{nt})\right]$$
$$= \mathbb{E}_\theta\left[(u(c_\tau^*(\theta)) - u(\bar{c}_\tau(\theta)))\big|x(I_{nt})\right]\cdot\mathbb{E}_\theta\left[\frac{1}{2}Tr\left(\frac{\partial P_{nt}}{\partial I_{nt}}\Sigma_\tau\right) - \frac{1}{2}\underbrace{(\theta-\mu_\tau)'\frac{\partial P_{nt}}{\partial I_{nt}}(\theta-\mu_\tau)}_{=Tr\left(\frac{\partial P_{nt}}{\partial I_{nt}}\Sigma_\tau\right)}\big|x(I_{nt})\right]$$
$$-\frac{1}{2}Cov\left(u(c_\tau^*(\theta)) - u(\bar{c}_\tau(\theta)), (\theta-\mu_\tau)'\frac{\partial P_{nt}}{\partial I_{nt}}(\theta-\mu_\tau)\big|x(I_{nt})\right)$$
$$= -\frac{1}{2}Cov\left(u(c_\tau^*(\theta)) - u(\bar{c}_\tau(\theta)), (\theta-\mu_\tau)'\frac{\partial P_{nt}}{\partial I_{nt}}(\theta-\mu_\tau)\big|x(I_{nt})\right)$$

Substituting,

$$\frac{\partial}{\partial I_{nt}}\bar{R}_\tau = -\frac{1}{2}\,\mathbb{E}_{\mathcal{I}_\tau,\nu_n,\nu_t}\left[\int_{\mathbb{R}^k} Cov\left(u(c^*_\tau(\theta)) - u(\bar{c}_\tau(\theta)), (\theta-\mu_\tau)'\frac{\partial P_{nt}}{\partial I_{nt}}(\theta-\mu_\tau)\big|x(I_{nt})\right) dx(I_{nt})\big|\mathcal{I}_t\right]$$

Applying the Law of Total Covariance,

$$\begin{aligned}
\frac{\partial}{\partial I_{nt}}\bar{R}_\tau &= -\frac{1}{2}\,\mathbb{E}_{\mathcal{I}_\tau,\nu_n,\nu_t,x}\left[-\frac{1}{2}Cov\left(u(c^*_\tau(\theta)) - u(\bar{c}_\tau(\theta)), (\theta-\mu_\tau)'\frac{\partial P_{nt}}{\partial I_{nt}}(\theta-\mu_\tau)\big|x(I_{nt})\right)\big|\mathcal{I}_t\right]\\
&= -\frac{1}{2}Cov\left(u(c^*_\tau(\theta)) - u(\bar{c}_\tau(\theta)), (\theta-\mu_\tau)'\frac{\partial P_{nt}}{\partial I_{nt}}(\theta-\mu_\tau)\big|\mathcal{I}_t\right)\\
&\quad -\frac{1}{2}Cov\left(\mathbb{E}_\theta\left[u(c^*_\tau(\theta)) - u(\bar{c}_\tau(\theta))|x(I_{nt})\right], \underbrace{\mathbb{E}_\theta\left[(\theta-\mu_\tau)'\frac{\partial P_{nt}}{\partial I_{nt}}(\theta-\mu_\tau)|x(I_{nt})\right]}_{=Tr(\frac{\partial P_{nt}}{\partial I_{nt}}\Sigma_\tau)}\big|\mathcal{I}_t\right)\\
&= -\frac{1}{2}Cov\left(u(c^*_\tau(\theta)) - u(\bar{c}_\tau(\theta)), (\theta-\mu_\tau)'\frac{\partial P_{nt}}{\partial I_{nt}}(\theta-\mu_\tau)\big|\mathcal{I}_t\right)
\end{aligned}$$

Therefore

$$\begin{aligned}
\frac{\partial}{\partial I_{nt}}\mathbb{E}[\bar{R}_{t+1}|\mathcal{I}_t] &= -\frac{1}{2}\sum_{\tau=t+2}^{\infty}\delta^{t-2}Cov\left(\left[u(c^*_\tau(\theta)) - u(\bar{c}_\tau(\theta))\right], (\theta-\mu_\tau)'\frac{\partial P_{nt}}{\partial I_{nt}}(\theta-\mu_\tau)\big|\mathcal{I}_t\right)\\
&= -\frac{1}{2}\sum_{\tau=t+2}^{\infty}\delta^{t-2}Cov\left(R_\tau(\theta), (\theta-\mu_\tau)'I_{nt}^{-2}V_{nt}^{-1}\Sigma_x V_{nt}^{-1}(\theta-\mu_\tau)\big|\mathcal{I}_t\right)
\end{aligned}$$

So the expected marginal reduction in regret associated with increasing I_{nt} is a function of the expected marginal reduction in the variance of posteriors times the covariance of regret and deviations of θ from its period τ expectation. This expression is typically negative since regret is minimized when $\theta = \mu_\tau$, in other words beliefs are correct.

This derivation also reveals several other intuitive features. $\frac{\partial}{\partial|\Sigma_x|}\frac{\partial\bar{R}}{\partial I_{nt}} > 0$, meaning that the marginal reduction in regret falls in magnitude if signals are less precise. Conversely, the value of information increases when learning from θ from sources other than I_{nt} become less precise. Since one must take an expectation over changes to the information set, this means that if φ increases so the agent expects to receive more information from neighbors, $\frac{\partial\bar{R}}{\partial I_{nt}}$ will fall in magnitude, reflecting the fact that the information is expected to be less useful. Similarly, if the agent has more precise priors, then information will have less value. Finally, if $R_\tau(\theta) = u(c^*_\tau(\theta)) - u(\bar{c}_\tau(\theta))$ falls, then $\frac{\partial\bar{R}}{\partial I_{nt}}$ will fall in magnitude because the utility cost of not knowing θ is lower, so information holds less value.

Proof of propositions

Proof of Proposition 1

I focus on the discrete choice of whether to stock at least χ units or not. This generates tractable predictions without requiring conditions on the smoothness of the hedging value of the mean-preserving contraction to ensure that it is differentiable with respect to I_{nt}. It also matches the mean-preserving contraction implemented in the experiment.

First note that the value of learning, $\mathbb{E}_t\left[\bar{R}(y_{t+1}, \mathcal{I}_{t+1})|I_{nt} = \chi\right]$, is unaffected by the mean-preserving contraction since it affects profit realizations only in period $t+1$, and regret is a function of periods beginning with $t+2$. The costs of stocking I_{nt} are also unaffected, and so the problem reduces to examining how the mean-preserving contraction affects the present value of expected utility associated with the contracted profits.

Applying Rothschild and Stiglitz (1970), there exists some random variable ϵ such that $\pi_n(\chi, \nu_{nt} + \theta) = \pi_n^p(\chi, \nu_{nt} + \theta) + \epsilon$ and satisfying $\mathbb{E}[\epsilon|\pi_n^p(\chi, \nu_{nt} + \theta)] = 0$.
Case 1: If the agent is risk neutral, then $u'(c_t) = \bar{u}$ is constant. The firm therefore is indifferent between redistributing profits across periods versus not, and so the present value of expected utility from investing $I_{nt} = \chi$ is proportional to $\delta\,\mathbb{E}_t[\pi_n^p(\chi, \nu_{nt} + \theta)]$ under the contraction and $\delta\,\mathbb{E}_t[\pi_n(\chi, \nu_{nt} + \theta)]$. By definition of the mean-preserving contraction, these values are the same and so the firm's utility from stocking $I_{nt} > 0$ is unaffected.[1]
Case 2: If the agent is risk averse, then $u''(c_t) < 0$. Consumption smoothing will lead the agent to borrow against future periods if profit realizations are low and save if they are high, but the present value of utility gains will remain a concave function of realized profits, which we may denote by $\varphi(\pi_n(\chi, \nu_{nt} + \theta))$ or $\varphi(\pi_n^p(\chi, \nu_{nt} + \theta))$.[2] By Jensen's Inequality and the Law of Iterated Expectations,

$$\begin{aligned}\mathbb{E}_t[\varphi(\pi_n(\chi, \nu_{nt} + \theta))] &= \mathbb{E}_t[\mathbb{E}_t[\varphi(\pi_n^p(\chi, \nu_{nt} + \theta) + \epsilon)|\pi_n^p(\chi, \nu_{nt} + \theta)]] \\ &< \mathbb{E}_t[\varphi(\mathbb{E}_t[\pi_n^p(\chi, \nu_{nt} + \theta) + \epsilon|\pi_n^p(\chi, \nu_{nt} + \theta)])] = \mathbb{E}_t[\varphi(\pi_n^p(\chi, \nu_{nt} + \theta))]\end{aligned}$$

Meaning that a risk averse agent gets strictly higher expected utility under the mean-preserving contraction, reflecting the fact that payouts are moved in towards the mean.
Case 3: If the agent is risk loving, the firm gets strictly lower expected utility under the mean-preserving contraction by a reverse argument to case 2.

Therefore the mean-preserving contraction increases the likelihood that $I^*_{nt} > 0$, in the sense that any agent that stocks under π_n will stock under π_n^p and there exists some agents (those with a particular distribution of ν_{nt}, θ) such that agents that do not stock under π_n stock under π_n^p, if and only if the agent is risk averse.

[1] This argument relies on the fact that risk neutral firms can always reduce consumption to reach the optimal point of investment. Absent this, a concave production function would yield a buffer stock savings problem that would induce risk averse behavior even with a constant marginal utility of consumption. This decision is based on the fact that firms empirically have relatively stationary monthly costs even if they have a negative shock, implying that they are able to reduce consumption to cover firm investments. Results would be similar if households had an exogenous flow of external income that they could invest in the business and profit functions were restricted such that losses do not exceed monthly consumption.

[2] If capital constraints are unbinding, then a Permanent Income Hypothesis result would imply that the agent's change in consumption is a fixed proportion of their change in profits. Capital constraints will lead to larger consumption reductions for low realizations.

It immediately follows that any variation $\pi_n^{p'}$ that is first order stochastic dominated by π_n^p will increase the likelihood that $I_{nt}^* > 0$ since all agents strictly prefer π_n^p.

Proof of Proposition 2

Part a: The result follows immediately for any increasing utility function since the realization of profits under returns first order stochastic dominates the profit function without returns.
Part b: An agent in period t_c will stock n if and only if the present value of expected utility gains from stocking it versus not exceed the utility loss of paying Γ. And for values where it is stocked, the present value of utility gains from stocking helmets is reducing in Γ since the agent must pay the restocking expense. Therefore regret is lower for any θ after this period since the possibly utility loss from not stocking n is lower, while regret is unaffected before that point. Thus, $\frac{\partial R_{t+1}(\theta)}{\partial \Gamma} < 0$. Therefore Equation 1.7 and 1.8 show that $\frac{\partial I_{nt}^*}{\partial \Gamma} < 0$ if and only if $|\Sigma_t| > 0$.
Part c: Capital constraints are not binding means that $\mathbb{E}_t \left[\frac{u'(c_{t+1})}{u'(c_t)}\right] = 1$ and $\zeta_n = 0 \Rightarrow \pi_n(\chi, \nu_{nt} + \theta) = \chi \cdot p_n(\chi, \nu_{nt+1}; \theta) \geq \chi \cdot w_n$.

Agents will therefore stock helmets if $Pr(\pi_n(\chi, \nu_{nt} + \theta) > 0) > 0$. This is immediate if agents are risk neutral. If they are risk averse, then if profit realizations of the safe good are high, the firm can save the consumption for the following period, ensuring that the utility of the consumption gains exceeds the foregone utility from stocking.

If $Pr(\pi_n(\chi, \nu_{nt} + \theta) > 0) = 0$, then regret is 0 and there is no learning value, so changes in Γ do not affect decisions and the agent never stocks. Therefore, if offered returns that eliminate loss risk, $\frac{\partial}{\partial \Gamma} I_{nt}^* = 0$.
Part d: If agents are risk neutral and have correctly centered beliefs, then expected profits from stocking χ are unaffected but learning value is lower so investment falls. If expected profits positively update or agents are sufficiently risk averse, then the expected utility of stocking increases so there will be persistent positive effects on stocking.

Proof of Proposition 3

Part i: Appendix A.1 shows that $\frac{\partial}{\partial |\Sigma_t|} \frac{\partial}{\partial I_{nt}} \mathbb{E}_t \left[\bar{R}_{t+1} | \mathcal{I}_t\right] < 0$, meaning that information has more value when priors are diffuse. Therefore investment is higher under more diffuse θ if the agent is risk neutral by Equation 1.7. The magnitude of $\frac{\partial}{\partial I_{nt}} \mathbb{E}_t \left[\bar{R}_{t+1} | \mathcal{I}_t\right]$ is falling as agents become more risk averse and the disutility of stocking increases. Therefore for a sufficiently risk averse agent, I_{nt}^* falls as uncertainty in θ increases.
Part ii: Returns do not affect the value of learning since regret is a function of payoffs beginning in period $t+2$. From Equation 1.7, it therefore follows that a risk neutral agent that stocks only with returns must have a lower $\mathbb{E}_t \left[\frac{u'(c_{t+1})}{u'(c_t)} \pi_n(\chi, \nu_{t+1} + \theta)\right]$ versus an agent that stocks without returns.

If the agent is risk averse, then returns also increases $Cov_t \left(u'(c_{t+1}), \frac{\partial}{\partial I_{nt}} \pi_{nt}(\chi, \nu_{nt+1} + \theta)\right)$, and so a firm that stocks only with returns may not have lower $\mathbb{E}_t \left[\frac{u'(c_{t+1})}{u'(c_t)} \pi_n(\chi, \nu_{t+1} + \theta)\right]$ than one that stocks without them if their priors are more diffuse, so that the increase in $Cov_t \left(u'(c_{t+1}), \frac{\partial}{\partial I_{nt}} \pi_n(\chi, \nu_{nt+1} + \theta)\right)$ is larger.

Part iii: Restocking shrinks $|\Sigma_t|$ which lowers in magnitude $\frac{\partial}{\partial I_{nt}} \mathbb{E}_t \left[\bar{R}_{t+1}|\mathcal{I}_t\right]$, meaning that learning has less value. By Equation 1.7, it follows that a risk neutral agent will obtain lower expected utility from stocking $I_{nt+1} = \chi$ unless $\mathbb{E}_{t+1}\left[\frac{u'(c_{t+1})}{u'(c_t)}\pi_n(\chi, \nu_{nt+2} + \theta)\right]$ increases since $u'(c)$ is constant so $Cov_t\left(u'(c_{t+2}), \frac{\partial}{\partial I_{nt+1}}\pi_n(I_{nt+1}, \nu_{nt+2} + \theta)\right) = 0$.

If agents are risk averse, then the contraction of beliefs about the profitability of helmets raises expected utility as proved earlier, so an agent may obtain higher expected utility from stocking $I_{nt+1} = \chi$ even if expected profits are unchanged or fall.

Proof of Proposition 4

Part a: Beginning with a change in Γ, a higher φ lowers $\bar{R}(\theta)$ since the agent expects to learn demand without entering themselves, lowering regret whenever helmets can be stocked. Therefore a change in Γ has a smaller impact since expected regret in period t_c is small under a high φ regardless of whether the agent can continue stocking n or not. Returns will also have lower effects since an agent with high φ has more precise beliefs about demand.

Parts b and c: The argument is similar to the prior proposition.

Proof that the insurance contract induces a mean-preserving contraction

First suppose that the premium payments to enterprises are

$$P_i = 1000 \cdot (1 - p_i)$$

and let the insurance contract be as given, paying out $1,000$ if sales are less than 3 and 0 if the shop sells 3 helmets. The payout P_i is strictly lower for all p_i compared to that used in the study, meaning the study version first order stochastic dominates it. Therefore proving that the version of the offer here is a mean-preserving contraction is sufficient to conclude that the insurance offer increases investment only if firms are risk averse.[3]

The restriction that shops cannot restock if they accept the insurance payout and the audits ensure that it is not profit increasing to intentionally sell fewer helmets or inflate prices after accepting the insurance contract, so I will assume that the agents' expected distribution of helmet sales is unaffected by opting into insurance. In particular, average profits conditional on selling 3 helmets exceed 800, so a firm capable of selling out always has higher expected profits by doing so than intentionally not and so has no incentive to follow a different sales strategy with insurance. Consistent with this assumption, prices were no higher on average among those that opted into insurance, those with insurance sold out at a higher rate than they anticipated, and all enterprises passed audits.

[3]The structure on the profit function differs from that imposed in the model, which requires strict monotonicity and smoothness conditions for tractability. Those conditions are imposed for model tractability (particularly with respect to learning), and neither is required in the proof of Proposition 1, so results are not sensitive to these differences.

Let $\pi_n(3, \nu_{nt+1} + \theta)$ be profits, inclusive of P_i, under the control offer and $\pi_n^p(3, \nu_{nt+1} + \theta)$ under the treatment offer. By construction, and recalling that $p_i = Pr(\text{sales}=3)$,

$$\begin{aligned}
\mathbb{E}_t[\pi_n^p(3, \nu_{nt+1} + \theta)] &= Pr(\text{sales} < 3) \cdot (1000 - P_i + \mathbb{E}_t[\pi_n(3, \nu_{nt+1} + \theta)|\text{sales} < 3]) \\
&\quad + Pr(\text{sales} = 3) \cdot (-P_i + \mathbb{E}_t[\pi_n(3, \nu_{nt+1} + \theta)|\text{sales} = 3]) \\
&= \underbrace{1000 \cdot (1 - p_i)}_{=P_i} - P_i + Pr(\text{sales} < 3) \cdot \mathbb{E}_t[\pi_n(3, \nu_{nt+1} + \theta)|\text{sales} < 3] \\
&\quad + Pr(\text{sales} = 3) \cdot \mathbb{E}_t[\pi_n(3, \nu_{nt+1} + \theta)|\text{sales} = 3] \\
&= \mathbb{E}_t[\pi_n(3, \nu_{nt+1} + \theta)]
\end{aligned}$$

where the last line leverages the Law of Total Probability. Therefore the expected profits are the same under the two offers. Assume that $p_i < 1$ since trivially the offers are the same if the agent perceives no risk of failing to sell out.

Let h_i be the price charged of a helmet. The proof requires that agents sell helmets for at least $1000 - P_i$, which is always true empirically. Let F denote the CDF of profits under π_n and F_p under π_n^p. Observe that F will make discrete jumps at P_i, $h_i + P_i$, $2h_i + P_i$ and $3h_i + P_i$ and F_p at 1000, $h_i + 1000$, $2h_i + 1000$ and $3h_i + 1000$ since payouts only change when demand crosses these thresholds.

For $x < 3 \cdot h_i$, $F_p(x) \leq F(x)$ and so we immediately have that $\int_{-\infty}^{x} F_p(y)dy \leq \int_{-\infty}^{x} F(y)dy$ and the inequality is strict at $x = 2 \cdot h_i + P_i$ since $p_i < 1$. For $y \geq 3h_i + P_i$, $F_p(y) = F(y) = 1$ and so if $\int_{-\infty}^{x} F_p(y)dy \leq \int_{-\infty}^{x} F(y)dy$ holds for $x \in [3h_i, 3h_i + P_i]$ we may conclude that π_p is a mean-preserving contraction. Observe that $F_p(y) = 1$ for all $y \in [3h_i, 3h_i + P_i]$ whereas $F(y) = 1 - p < 1$. Therefore it suffices to verify that $\int_{-\infty}^{3h_i+P_i} F_p(y)dy \leq \int_{-\infty}^{3h_i+P_i} F(y)dy$.

Let $\mathbb{P}_0 = Pr(\text{sales} = 0)$, $\mathbb{P}_1 = Pr(\text{sales} = 1)$ and $\mathbb{P}_2 = Pr(\text{sales} = 2)$.

$$\begin{aligned}
\int_{-\infty}^{3h_i+P_i} F(y)dy &= \int_{-\infty}^{P_i} F(y)dy + \int_{P_i}^{P_i+h_i} F(y)dy + \int_{P_i+h_i}^{P_i+2h_i} F(y)dy + \int_{P_i+2h_i}^{P_i+3h_i} F(y)dy \\
&= h_i \cdot \mathbb{P}_0 + h_i \cdot (\mathbb{P}_0 + \mathbb{P}_1) + h_i \cdot (\mathbb{P}_0 + \mathbb{P}_1 + \mathbb{P}_2) \\
&= h_i \cdot (3\mathbb{P}_0 + 2\mathbb{P}_1 + \mathbb{P}_2)
\end{aligned}$$

Case 1: Suppose that $h_i \geq 1,000$.

$$\begin{aligned}
\int_{-\infty}^{3h_i+P_i} F_p(y)dy &= \int_{-\infty}^{1000} F_p(y)dy + \int_{1000}^{1000+h_i} F_p(y)dy + \int_{1000+h_i}^{1000+2h_i} F_p(y)dy \\
&\quad + \int_{1000+2h_i}^{3h_i} F(y)dy + \int_{3h_i}^{3h_i+P_i} F(y)dy \\
&= h_i \cdot \mathbb{P}_0 + h_i \cdot (\mathbb{P}_0 + \mathbb{P}_1) + (h_i - 1000) \cdot (\mathbb{P}_0 + \mathbb{P}_1 + \mathbb{P}_2) \\
&\quad + \underbrace{P_i}_{=1000 \cdot (\mathbb{P}_0 + \mathbb{P}_1 + \mathbb{P}_2)} \cdot (\mathbb{P}_0 + \mathbb{P}_1 + \mathbb{P}_2 + \mathbb{P}_3) \\
&= h_i \cdot (3\mathbb{P}_0 + 2\mathbb{P}_1 + \mathbb{P}_2)
\end{aligned}$$

Case 2: Suppose that $1000 - P_i \leq h_i < 1000$.

$$\begin{aligned}\int_{-\infty}^{3h_i+P_i} F_p(y)dy &= \int_{-\infty}^{1000} F_p(y)dy + \int_{1000}^{1000+h_i} F_p(y)dy + \int_{1000+h_i}^{3h_i} F_p(y)dy + \int_{3h_i}^{2h_i+1000} F(y)dy \\ &\quad + \int_{2h_i+1000}^{3h_i+P_i} F(y)dy \\ &= \int_{1000+h_i}^{3h_i} F_p(y)dy + \int_{3h_i}^{2h_i+1000} F(y)dy + \int_{3h_i}^{3h_i+P_i} dy - \int_{3h_i}^{2h_i+1000} dy \\ &= h_i \cdot \mathbb{P}_0 + (2h_i - 1000) \cdot (\mathbb{P}_0 + \mathbb{P}_1) + (1000 - h_i) \cdot (\mathbb{P}_0 + \mathbb{P}_1 + \mathbb{P}_3) \\ &\quad + P_i - (1000 - h_i) \\ &= h_i \cdot (3\mathbb{P}_0 + 2\mathbb{P}_1 + \mathbb{P}_2)\end{aligned}$$

Therefore the insurance contract is a mean-preserving contraction.

A.2 Coefficient of relative risk aversion estimation

This section estimates the coefficient of relative risk aversion of firms' implied by the effects of the contract offered in the Insurance Experiment. Agents indexed by i face a discrete choice to stock 3 helmets, $s_i = 1$, or not $s_i = 0$. Utility from profits is given by the utility function

$$u(\pi_i) = \begin{cases} \frac{\pi_i^{1-\theta}}{1-\theta}, & \theta \neq 1 \\ \log(\pi_i), & \theta = 1 \end{cases}$$

where θ is the coefficient of relative risk aversion. Absent insurance, agents' expected utility from stocking is given by

$$\begin{aligned}\mathbb{E}[u(\pi_i^{noins})] &= \gamma + u(P_i) \cdot Pr(Sales_i = 0) + u(price_i + P_i) \cdot Pr(Sales_i = 1) \\ &\quad + u(2 \cdot price_i + P_i) \cdot Pr(Sales_i = 2) + u(3 \cdot price_i + P_i) \cdot Pr(Sales_i = 3)\end{aligned}$$

where γ captures utility or disuility from stocking helmets not related to their expected profits, in particular capital constraints or a hassle cost of learning to stock. P_i denotes the insurance premium offered to the firm and $price_i$ denotes the price firms charge per helmet.

Treated firms have access to insurance. Stocking with insurance yields expected utility of profits

$$\begin{aligned}\mathbb{E}[u(\pi_i^{\text{ins}})] &= \gamma + u(1000) \cdot Pr(Sales_i = 0) + u(price_i + 1000) \cdot Pr(Sales_i = 1) \\ &\quad + u(2 \cdot P_i + 1000) \cdot Pr(Sales_i = 2) + u(3 \cdot P_i) \cdot Pr(Sales_i = 2)\end{aligned}$$

Let z_i denote treatment assignment. Assuming agents' optimally opt into insurance when offered,

$$\mathbb{E}[u(\pi_i^{stock}|z_i)] = \begin{cases} \max\{\mathbb{E}[u(\pi_i^{\text{noins}})], E[u(\pi_i^{\text{ins}})]\} + \epsilon_{is}, & z_i = 1 \\ \mathbb{E}[u(\pi_i^{\text{noins}})] + \epsilon_{is}, & z_i = 0 \end{cases}$$

where $\epsilon_{ih} \sim EV1$ denotes firm-specific determinants of utility from stocking unobserved to the econometrician. Absent stocking, agents retain the investment cost and receive the insurance premium. Therefore

$$\mathbb{E}[u(\pi_i^{nostock})] = u(2210 + P_i) + \epsilon_{in}$$

where $\epsilon_{in} \sim EV1$. A firm's probability of stocking is therefore given by

$$Pr(s_i = 1|z_i) = \frac{\exp(u(\pi_i^{stock}|z_i))}{\exp(u(\pi_i^{stock}|z_i)) + \exp(u(\pi_i^{nostock}))}$$

Probabilities of selling each number of helmets and prices are observed. The aim is to identify the coefficient of relative risk aversion, θ, and γ, which absorbs features such as liquidity constraints. I leverage the moment conditions $\mathbb{E}[s_i - Pr(s_i = 1|z_i)] = 0$ and $\mathbb{E}[z_i(s_i - Pr(s_i = 1|z_i))] = 0$ where the second moment condition holds by random assignment of insurance, assuming that insurance only affects decisions via expected payouts.

The model is estimated in Python using differential evolution, searching over $\theta \in [0, 8]$ and $\gamma \in [-25, 25]$. The estimated value of θ is 0.62, although bootstrapped confidence intervals are uninformaive since the model is non-linear and estimated with only 350 observations. I also estimate the model for firms exhibiting below or above-median risk aversion measured via lottery choice, and find that the less informative group has an estimated $\theta = 0.53$ and the more risk averse $\theta = 2.21$. Although suggestive, these estimates align well with the heterogeneity from the lottery choice measures of risk aversion. This suggests that even modest levels of risk aversion can substantially affect firms' new product stocking.

Appendix B

Appendix to Chapter 2

B.1 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.2 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 2.3. Recall that the two-stage least squares model which identifies VSL is given by

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

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

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

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

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

Keeping $r_{0,i}$ fixed, 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 B.5. 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.

B.3 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. The primary VSL estimates using a measure of r_{i0} distinct from that in Δr_i to avoid attenuation from measurement error passed through to the instrument. This may lead to a LATE that weights observations between these stylized limit cases, but the instrument still weights observations more heavily in priors so the logic persists. Furthermore, the test results hold using the same measure of r_{i0} used to construct Δr_i.

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, $\mathbb{E}[\zeta_i r_{i0}] \neq 0$.

B.4 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. Liu et al. (2008) main inclusion criteria is that studies have a clear control group: the authors note that most papers fail to account for selection, and a large share use data only on those that end up hospitalized, which fails to account for victims that never enter a hospital. This places Ouellet and Kasantikul (2006), which attempts to reach all accident scenes rather than leveraging data on those admitted to hospitals, among the higher quality group of studies, although it still includes minimal corrections for selection into helmet usage (e.g. by safer drivers). In short, there are limitations to both Liu et al. (2008) and Ouellet and Kasantikul (2006) such that neither is clearly dominant, yet both are in line with the quality of evidence that practitioners have recently used in health campaigns.

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.

B.5 Appendix Figures

Figure B.1: Survey locations

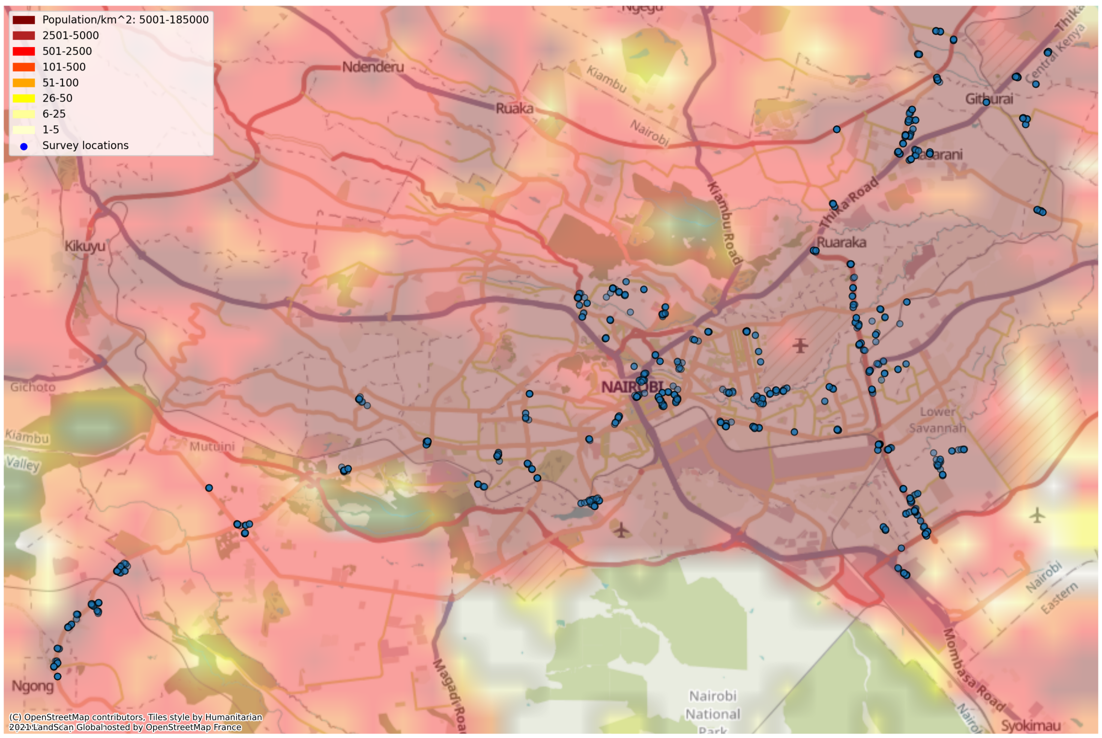
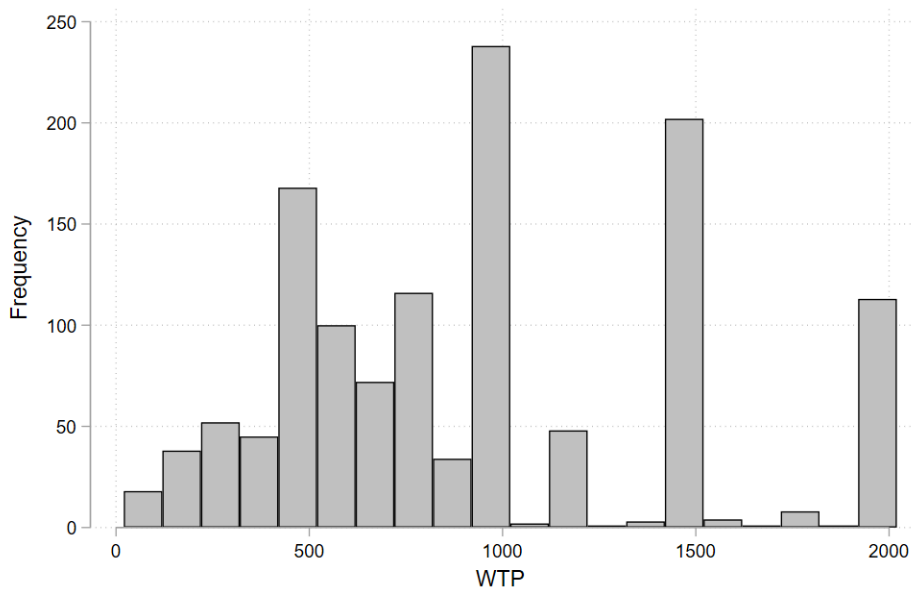


Figure B.1 plots survey locations over a map of Nairobi. Color denotes population, where no color denotes no residents. The 2019 census estimated a population of about 4.4 million within the city. Map data is from Open Street Map. Population data is from the LandScan Global 2021 data set produced by Oak Ridge National Laboratory.

Figure B.2: Distribution of helmet bids (Kenyan shillings)

(a) A. Histogram of bids, restricted axis



(b) B. Histogram of bids, full distribution

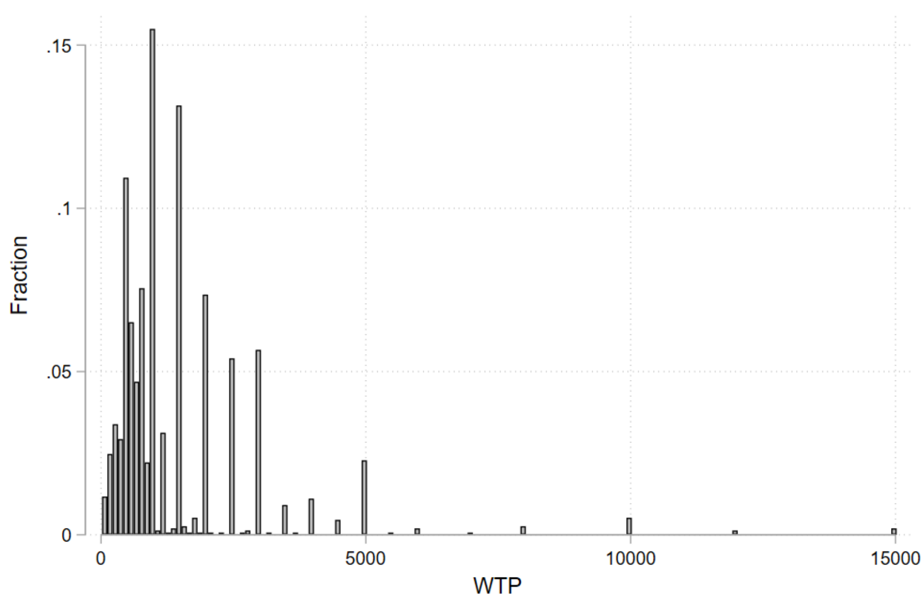


Figure B.2 plots the distribution of willingness to pay for helmets in Kenyan shillings (Ksh). Panel A reports a histogram of bids, excluding outliers above Ksh 2,000 for clarity. Panel B reports the same histogram across the full distribution of bids. Figures present Ksh rather than USD to illustrate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

Figure B.3: Effect of study VSL on published benefit-cost ratios

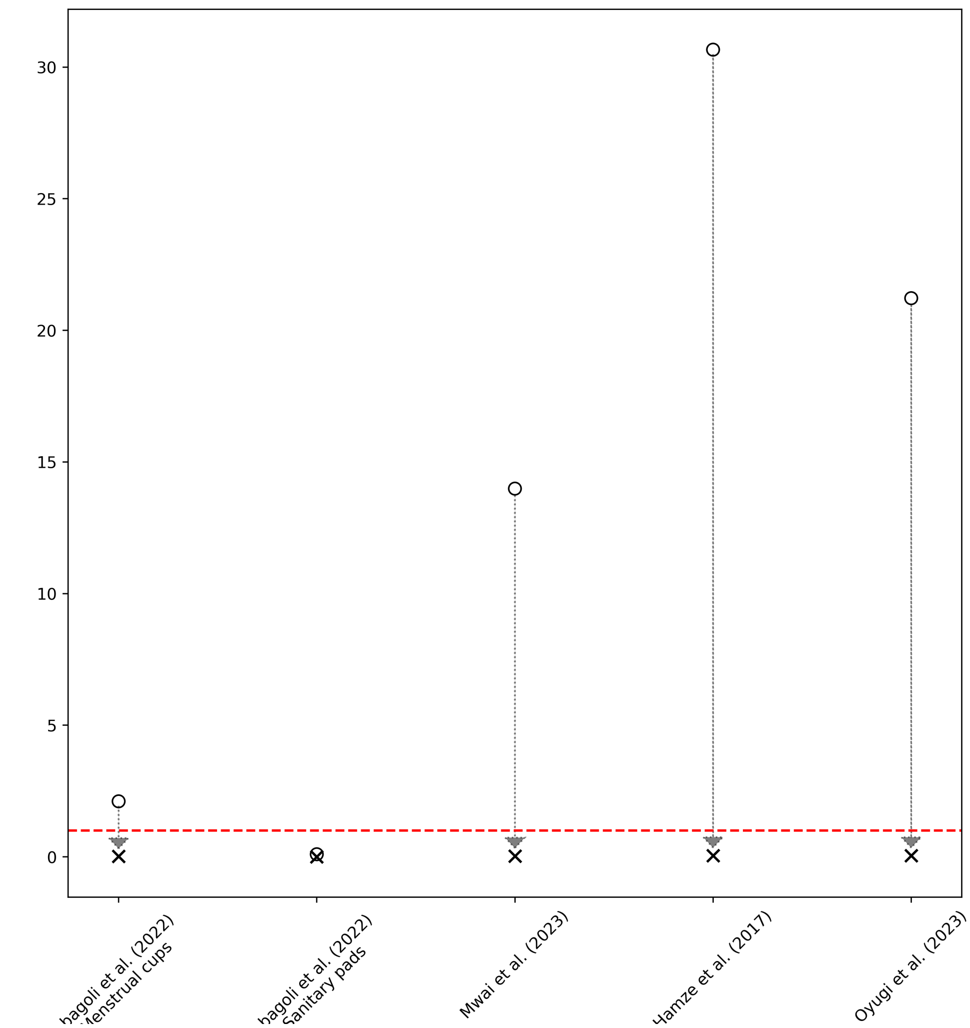
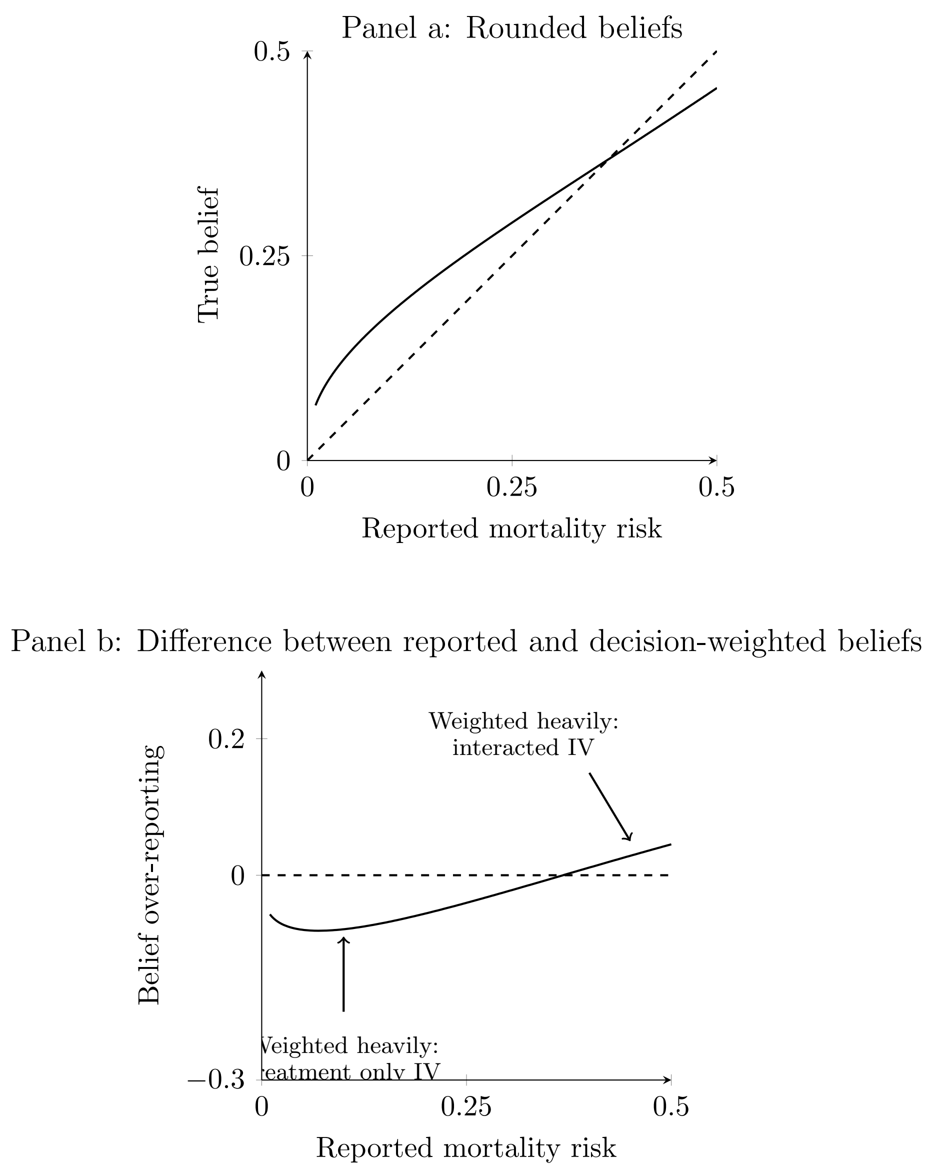


Figure B.3 examines how the benefit-cost ratios (BCRs) of published benefit-cost analyses of Kenyan programs change when the study's original VSL or VSLY is replaced with the preferred estimates from this study. Hollow circles denote the original BCR estimate, and an x denotes the revised estimate. The first two estimates are from Babagoli et al. (2022). The third estimate is from Mwai et al. (2023), the fourth is from Hamze, Mengiste, and Carter (2017), and the final figure is from Oyugi et al. (2023). The horizontal red line is at $BCR = 1$, the threshold for benefits exceeding costs.

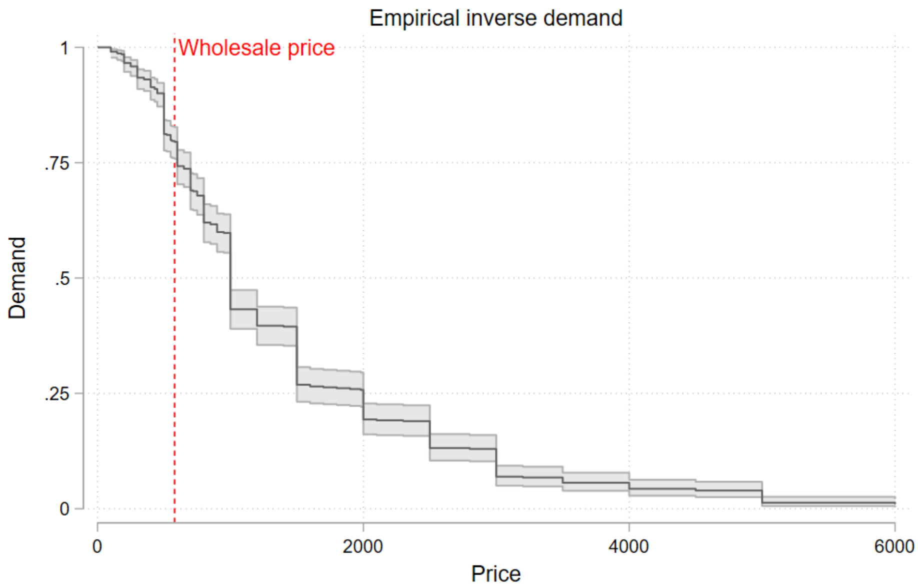
Figure B.4: Belief distortion under rounding



Appendix Figure B.4 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 overestimates VSL versus the "treatment only" instruments.

Figure B.5: Demand for helmets

(a) Inverse demand, control group



(b) Elasticity of demand, control group

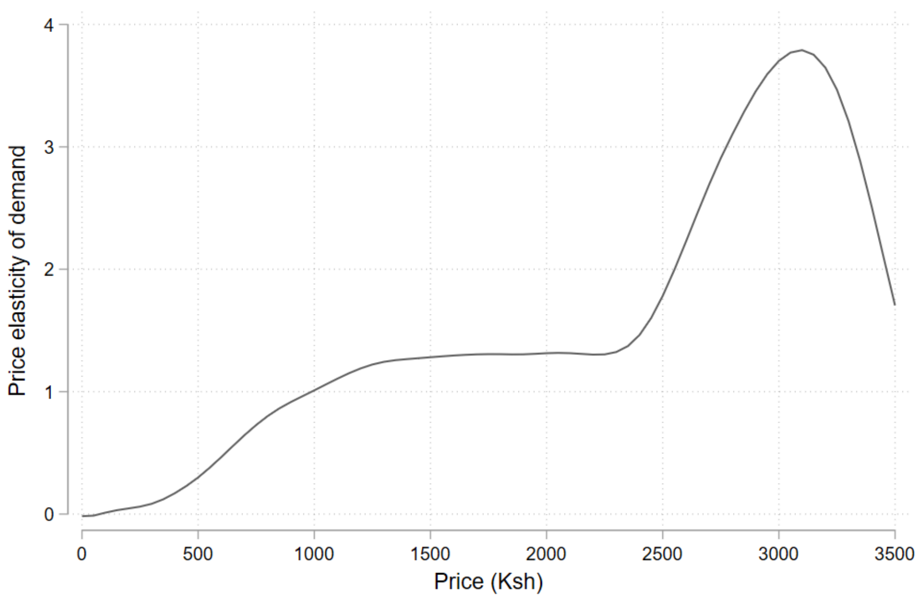


Figure B.5 plots estimates of demand for helmets among observations in the control and pure control arms. The plot of demand includes a pointwise confidence interval. The plot of the elasticity of demand is based on a local polynomial estimation adapted from Berry, Fischer, and Guiteras (2020). The vertical 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.

B.6 Appendix Tables

Table B.1: 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.

Table B.2: Elasticity of helmet demand with respect to subjective beliefs

	All observations		Information control	
	(1) Motorcycle risk	(2) Motorcycle risk	(3) Helmet effectiveness	(4) Helmet effectiveness
Elasticity	0.012	0.014	0.420	0.441
	(0.007)	(0.007)	(0.161)	(0.159)
Controls	No	Yes	No	Yes
Observations	1,421	1,421	452	452

Robust standard errors in parenthesis.

Table B.2 reports the elasticity of willingness to pay for a helmet with respect to priors about the likelihood of suffering a fatal accident in columns (1) - (2) and perceived efficacy of a helmet at preventing death among the information control arm in columns (3) - (4). Columns (2) and (4) include controls for the log of income, digit span recall, and log years of education. Estimates exclude 35 observations for which surveyors reported motorcycle taxi drivers pretending to be passengers.

Table B.3: Value of a statistical life: Winsorized beliefs

	Raw WTP				Winsorized WTP	
	(1) Interacted	(2) Treatment only	(3) Interacted	(4) Treatment only	(5) Interacted	(6) Treatment only
VSL	1,022.04	671.97	1,054.78	653.89	712.40	436.77
	(504.14)	(391.26)	(522.87)	(426.06)	(374.03)	(324.33)
Cragg-Donald F-stat	4.89	8.98	4.92	7.78	4.92	7.78
Weak IV Robust Confidence Set	[163.69, 2,399.38]	[-56.16, 1,833.88]	[163.69, 2,399.38]	[-172.73, 2,054.10]	[75.57, 1,526.95]	[-166.81, 1,374.25]
Inversion test	CLR	CLR	CLR	CLR	CLR	CLR
Observations	1,425	1,425	1,425	1,425	1,425	1,425
Controls	BL Risk	BL Risk	LASSO	LASSO	LASSO	LASSO
Enumerator FE	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parenthesis.

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.

Table B.4: Robustness of VSL to alternative assumptions

Panel A: Change in planned future ridership

	Raw WTP				Winsorized WTP	
	(1) Interacted	(2) Treat only	(3) Interacted	(4) Treat only	(5) Interacted	(6) Treat only
VSL	307.83	287.46	329.53	264.60	223.82	199.55
	(132.52)	(187.29)	(130.79)	(195.50)	(96.08)	(144.70)
F-stat	10.90	10.37	11.24	9.52	11.21	9.52
Weak IV Conf. Set	[75.40, 716.79]	[-82.08, 803.62]	[104.96, 730.19]	[-130.39, 811.47]	[54.46, 502.22]	[-86.79, 578.53]
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) Treat only	(3) Interacted	(4) Treat only	(5) Interacted	(6) Treat only
VSL	469.59	332.13	491.65	298.61	347.24	240.87
	(162.98)	(245.11)	(162.87)	(246.66)	(118.40)	(182.05)
F-stat	13.03	10.38	13.07	10.06	13.07	10.06
Weak IV Conf. Set	[195.82, 939.79]	[-151.15, 1,032.40]	[220.06, 960.00]	[-198.32, 988.80]	[141.71, 666.78]	[-118.93, 712.36]
Observations	1,411	1,425	1,411	1,425	1,411	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) Treat only	(3) Interacted	(4) Treat only	(5) Interacted	(6) Treat only
VSL	361.00	96.56	366.01	86.40	234.72	110.97
	(146.50)	(311.00)	(146.47)	(304.39)	(107.54)	(225.65)
F-stat	20.12	7.97	19.81	8.26	19.81	8.26
Weak IV Conf. Set	[82.68, 691.70]	[-637.11, 865.77]	[88.88, 699.07]	[-632.16, 837.48]	[27.21, 470.51]	[-388.60, 671.41]
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.

Table B.5: VSL: GMM and IV probit estimates with robust standard errors

	GMM: Raw WTP		GMM: Wins. WTP		IV probit	
	(1)	(2)	(3)	(4)	(5)	(6)
	Interacted	Treat only	Interacted	Treat only	Treat only	Treat only
VSL	464.03	308.12	353.47	252.59	161.25	161.92
	(228.88)	(238.51)	(178.48)	(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)
F-stat	6.18	10.96	6.18	10.96	NA	NA
Weak IV	[1.85,	[-154.63,	[7.19,	[-122.64,	NA	NA
Conf. Set	1,306.84]	978.63]	883.49]	750.35]		
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

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

Table B.5 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.

Table B.6: Correlates between perceived mortality risk and demographics

	(1) Years of education	(2) Female	(3) log(Income)	(4) Age	(5) Above-median health	(6) 1(Children)
Coefficient	2.20	-95.56	43.83	-1.20	-57.59	53.09
	(8.15)	(49.46)	(29.20)	(2.37)	(46.69)	(52.22)
Sample average	333.22	333.22	333.22	333.22	333.22	
Observations	1,419	1,427	1,221	1,425	1,427	1,427

Robust standard errors in parenthesis.

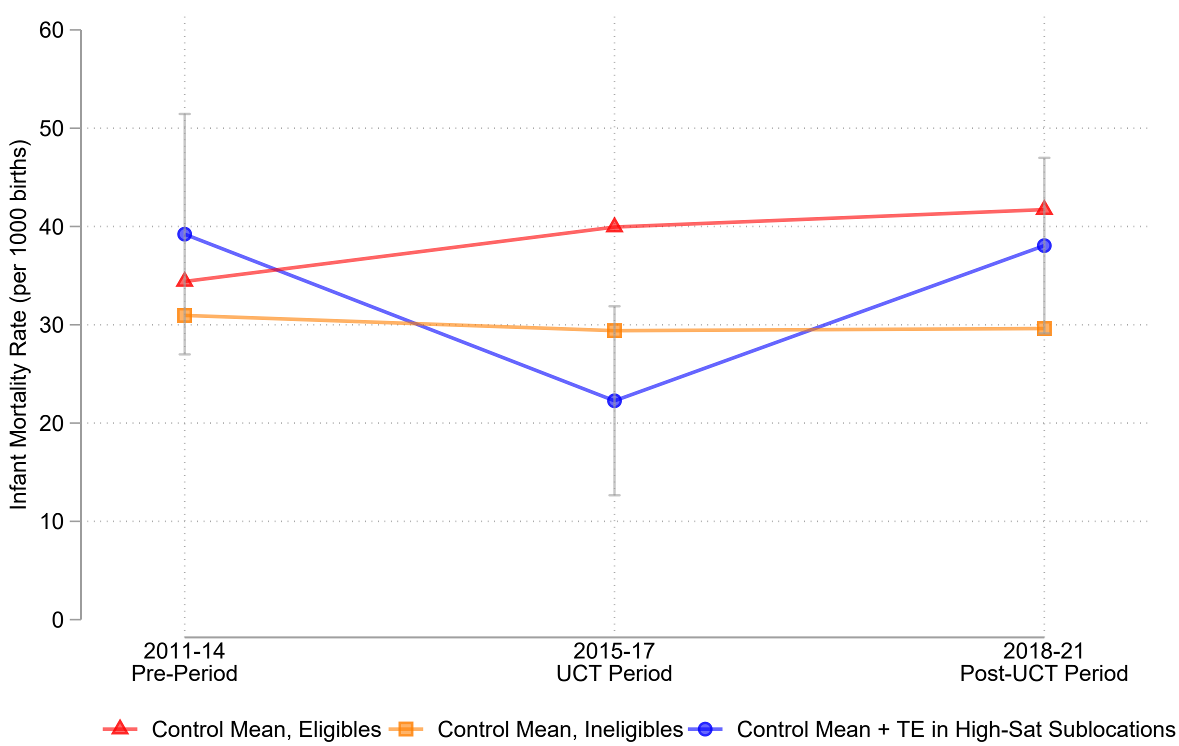
Table B.6 reports coefficients of regressions of perceived baseline mortality risk on demographic covariates. As detailed in the results, the local average VSL identified by this study weights observations with higher perceived risk more heavily, so this table aims to understand whether the LATE is over a selected population. Column (1) considers years of education, (2) an indicator equal to 1 if the respondent is female, (3) the log of income, (4) respondent age, (5) an indicator for reporting above-median health, and (6) an indicator for having children. No controls are colluded since the aim is only to understand what covariates are correlated with perceived risk. Estimates exclude 35 observations for which surveyors reported motorcycle taxi drivers pretending to be passengers.

Appendix C

Appendix to Chapter 3

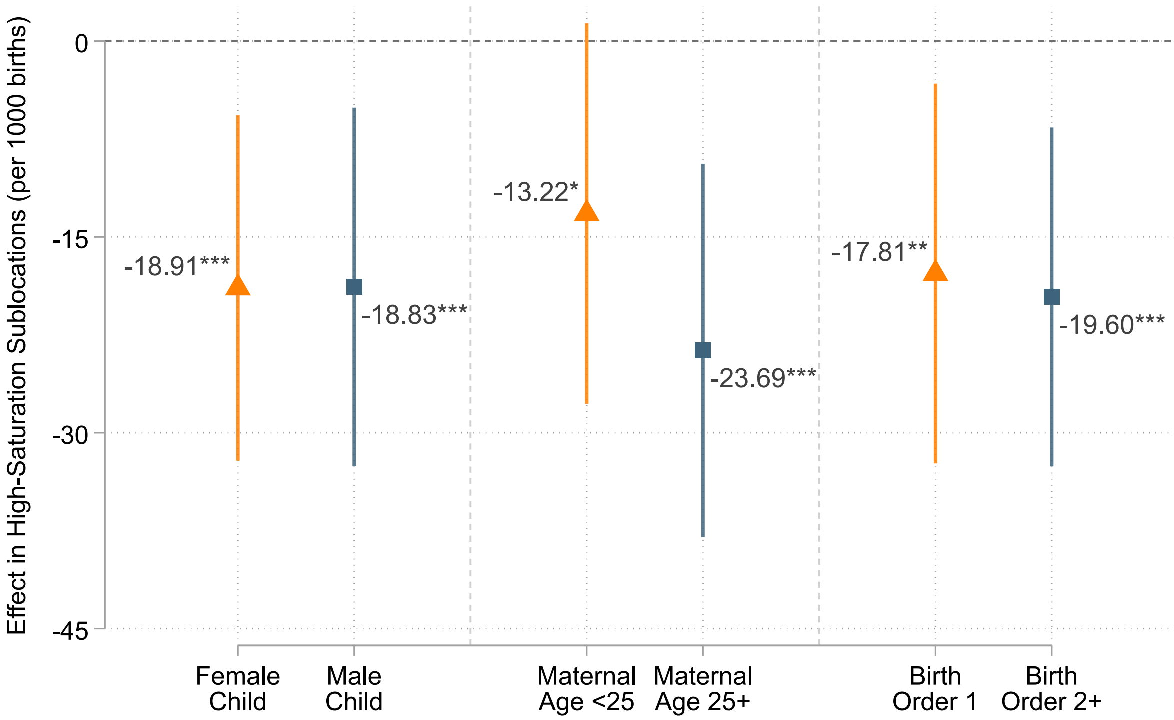
C.1 Appendix Tables and Figures

Figure C.1: Unconditional Cash Transfers and Infant Mortality



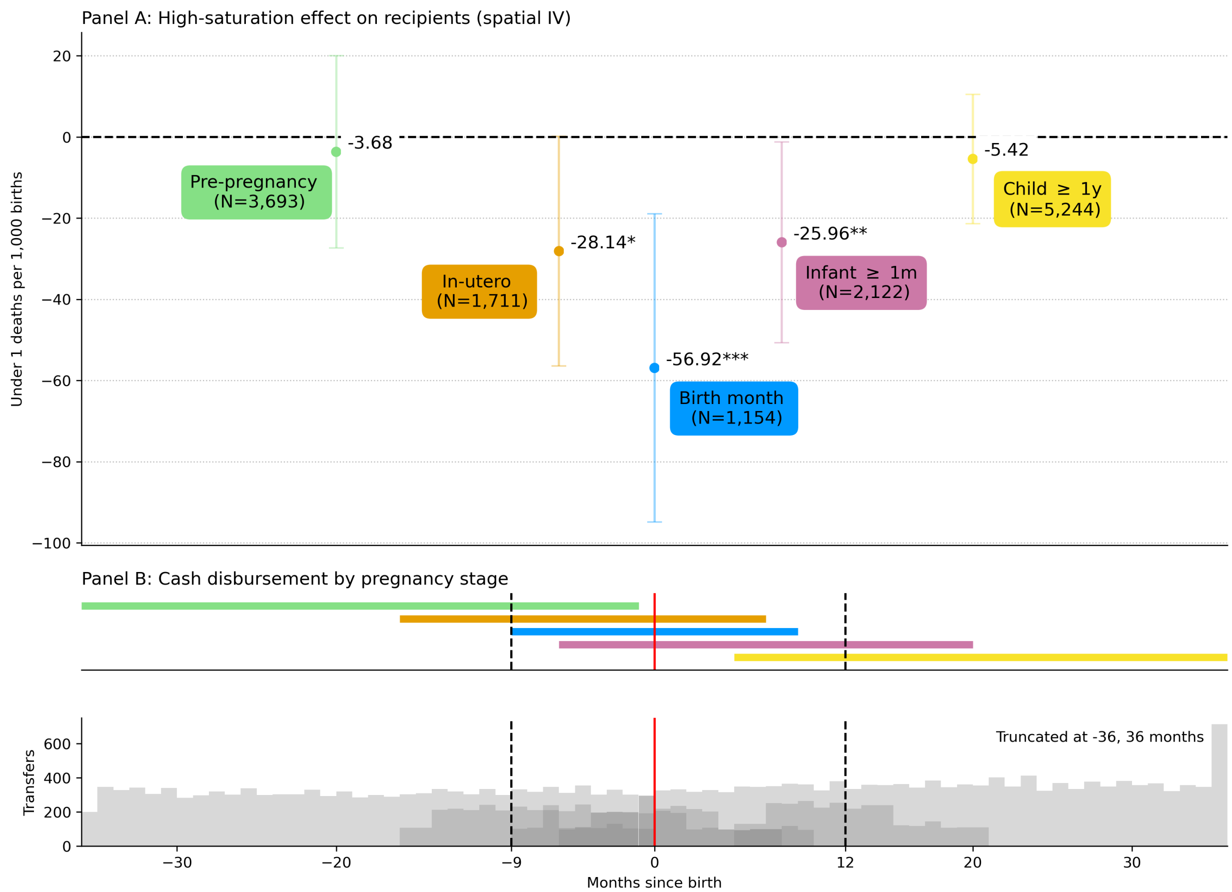
Notes: This figure is based on the main EL3 birth census sample encompassing births from 2011 to 2021. The figure plots the infant (under-1) mortality rate per 1000 births (y-axis) for three different periods of child births: pre-transfers (child birth years of 2011-14), the transfer period (child birth years 2015-17), and the post-transfer period (child birth years 2018-21). The red line reports average rates for transfer-eligible households in control, low saturation villages. The blue line adds in the estimated treatment effects for treatment villages in high saturation areas by period (from Equation (3.1), reported in Table 3.1), with 95% confidence intervals shown. The yellow line reports infant mortality rates for transfer-ineligible households residing in control, low saturation villages.

Figure C.2: Heterogeneity by Gender, Maternal Age, and Birth Order



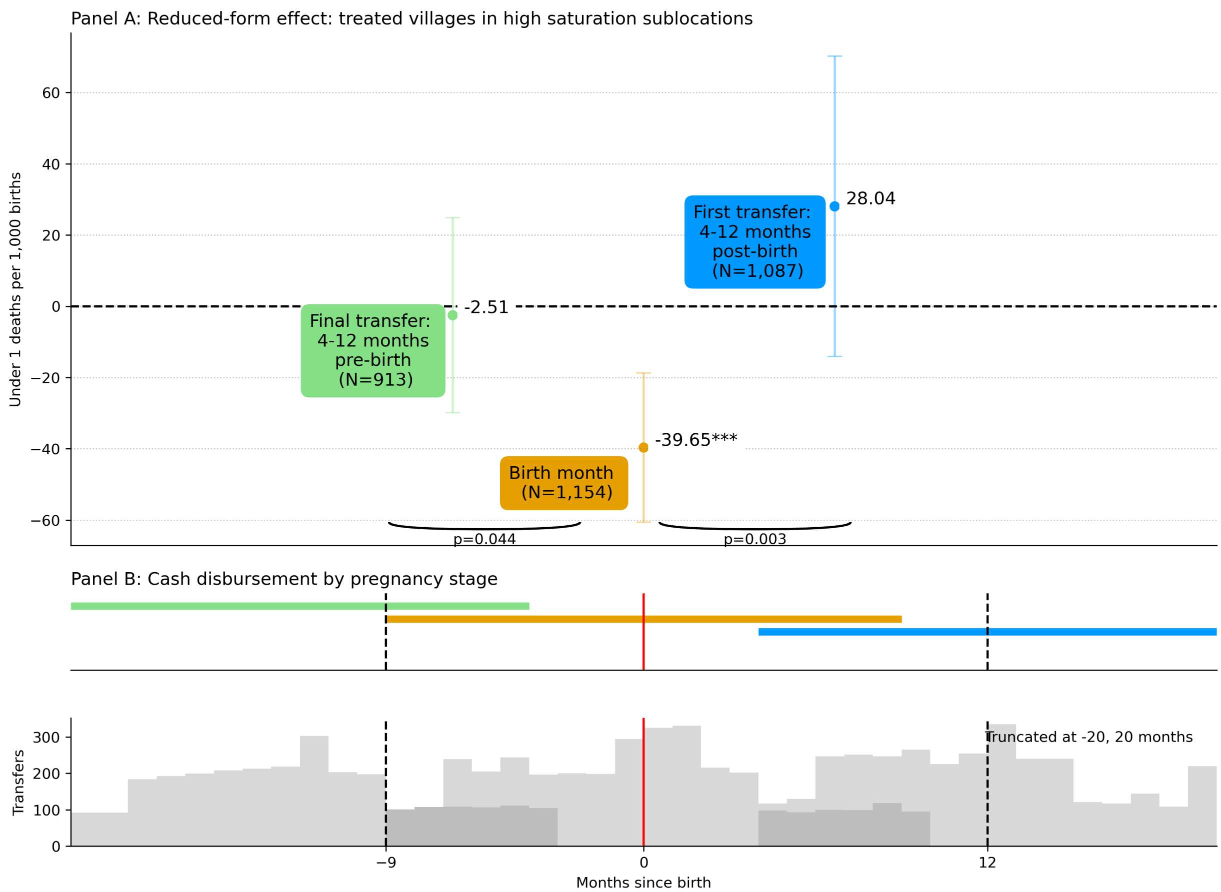
Notes: This figure is based on the main EL3 birth census sample, and plots estimated effects of cash transfers on infant mortality in treatment villages within high-saturation sublocations by gender, maternal age, and birth order. Maternal age at birth is analysed using pre-specified five-year age groups (under 20, 20-24, 25-29, 30-34, and 35 and above); this analysis splits on the median five-year group, with 46% of births having a mother below the age of 25. Child birth order is calculated using the order of births across the 2011-23 period and does not take into account births to mothers prior to 2011 as information on those births was not obtained in the census. Effects are estimated using equation (3.1) augmented with an interaction with the characteristic of interest. The p-values for the difference across groups (the effect on the interaction term in treatment villages in high-saturation sublocations) is as follows: 0.993 (gender), 0.337 (maternal age), and 0.856 (birth order). The results shown in this figure are additionally presented in Appendix Table C.8. Standard errors are clustered at the sublocation level. 95% confidence intervals are shown. * $p < .10$, ** $p < .05$, *** $p < .01$.

Figure C.3: Transfer Effects on Infant Mortality by Timing of Cash (Spatial IV)



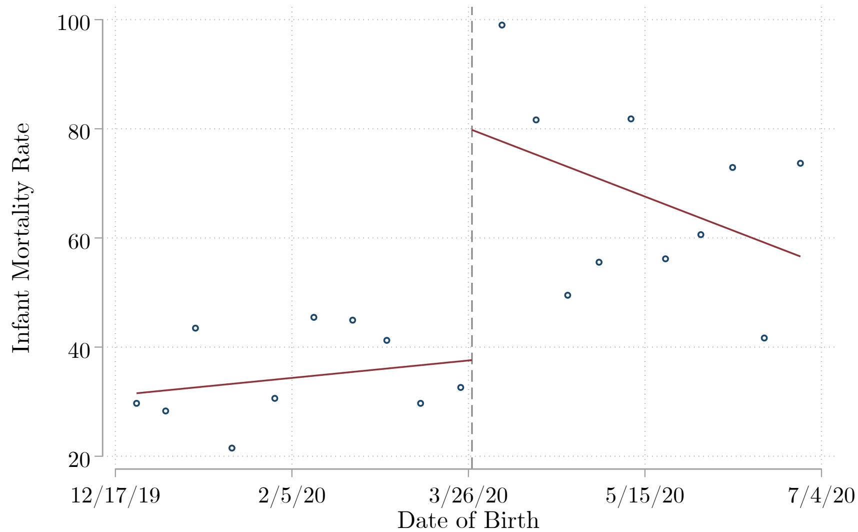
Notes: This figure is based on the main EL3 birth census sample. Panel A plots estimates of dynamics based on the time period of exposure to cash during the child's life. The transfer timing is defined relative to the "experimental start date", as this is well-defined for both treatment and control villages. "Pre-pregnancy" includes household exposure 3 years to 10 months before birth. In-utero is 9 to 1 month before birth. Birth month includes cash within the first month of life. Infant includes 1 month to 12 months. Child includes 1 to 3 years. Estimates are constructed using the same approach as equation (3.4), but using the spatial IV. We estimate equation (3.2) after restricting the sample to those exposed to cash at a particular time in life. Observations appear in multiple groups since cash transfers were distributed over 8 months. Panel B plots the range of experimental start dates, relative to birth month, included in each estimate and a histogram of transfers by month in each bin. Since the transfers were distributed in a series of three payments, some households in the 1-2 years post-transfer bin were still receiving cash transfers during this period. The reduced-form version of this figure (which is estimated using equation 3.1) is presented as Figure 3.2. 95% confidence intervals are shown. * $p < .10$, ** $p < .05$, *** $p < .01$.

Figure C.4: Effects of Cash in Birth Month vs Neighboring Times on Mortality (Reduced-form)

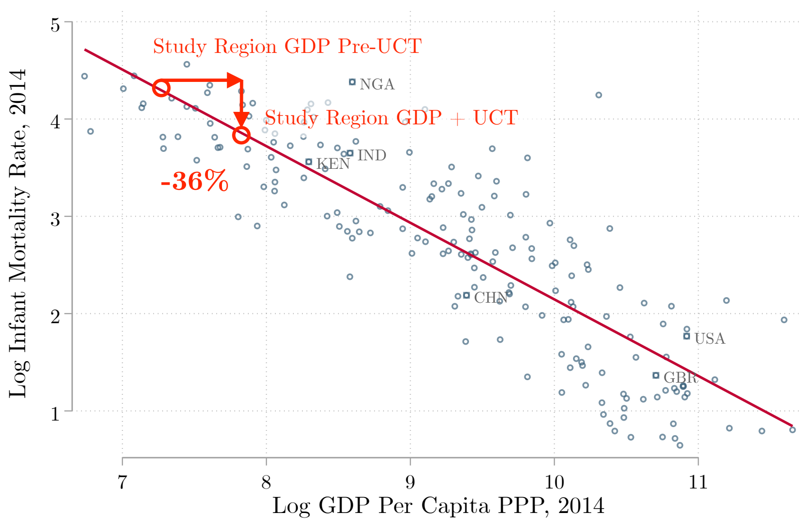


Notes: This figure is based on the main EL3 birth census sample. Panel A plots estimates of dynamics based on the time period of exposure to cash during the child's life. The transfer timing is defined relative to the "experimental start date", as this is well-defined for both treatment and control villages. The first group includes households where the last transfer was received between 12 and 4 months before birth (or would have been if they are control). Birth month includes cash within the first month of life. The third group includes birth where the experimental start date was 4-12 months post-birth. Estimates are constructed using the same approach as equation (3.4). We estimate equation (3.1) after restricting the sample to those exposed to cash at a particular time in life. The bottom of Panel A reports p-values for equality to the birth month effect based on an estimate which includes both groups indicates, and interacts the treatment effects with cash receipt timing. Panel B plots the range of experimental start dates, relative to birth month, included in each estimate and a histogram of transfers by month in each bin. 95% confidence intervals are shown. * $p < .10$, ** $p < .05$, *** $p < .01$.

Figure C.5: Non-Experimental Variation in Child Mortality



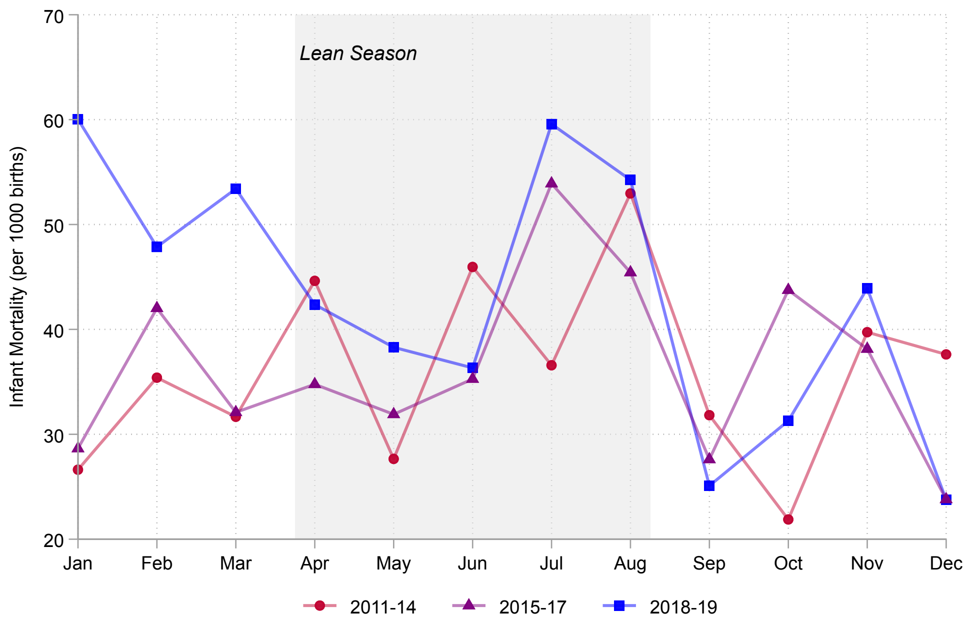
Panel A: COVID-19 Lockdown and Infant Mortality in Study Region



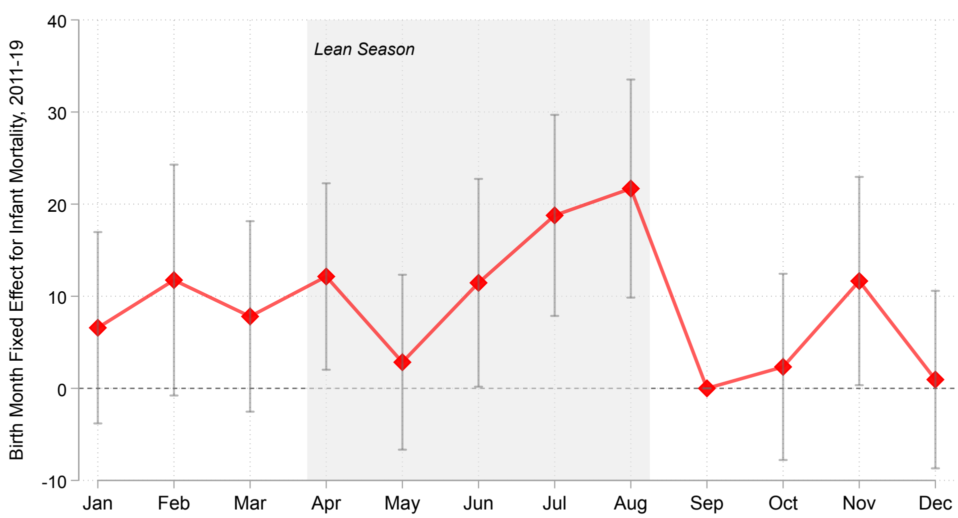
Panel B: Cross-Country GDP - Infant Mortality Relationship

Notes: This figure is based on the main EL3 birth census sample. Panel A reports infant mortality in the study region before and after Kenya's March 2020 COVID-19 lockdown using exact date of birth data from this project's birth census, utilizing all censused births with available data. The data in Panel B are shown as a binned scatter plot. Panel B reports the cross-country association between the natural log of the infant mortality rate (per 1000 births) as reported by the United Nations, 2024 and the natural log of GDP per capita (in USD PPP) in 2014 as reported by the World Bank, 2024. The "Study Region GDP Pre-UCT" dot highlights the study region's per capita GDP in 2014 as estimated by the Kenya National Bureau of Statistics, 2019 as well as predicted infant mortality based on the cross-country relationship. The "Study Region GDP + UCT" dot illustrates the study region's per capita GDP adding a UCT of the same fraction of GDP as the transfer disbursed as a fraction of household consumption (75%) as part of this project (and – consistent with the evidence – assuming it is spent within a year), and the predicted change in infant mortality at that level of income based on the cross-country relationship. Note that the Kenyan national average ("KEN") lies nearly on the regression line and at a higher per capita income level than the rural study region, since the national average also include far richer urban areas like Nairobi.

Figure C.6: Seasonality in Infant Mortality in the Birth Census



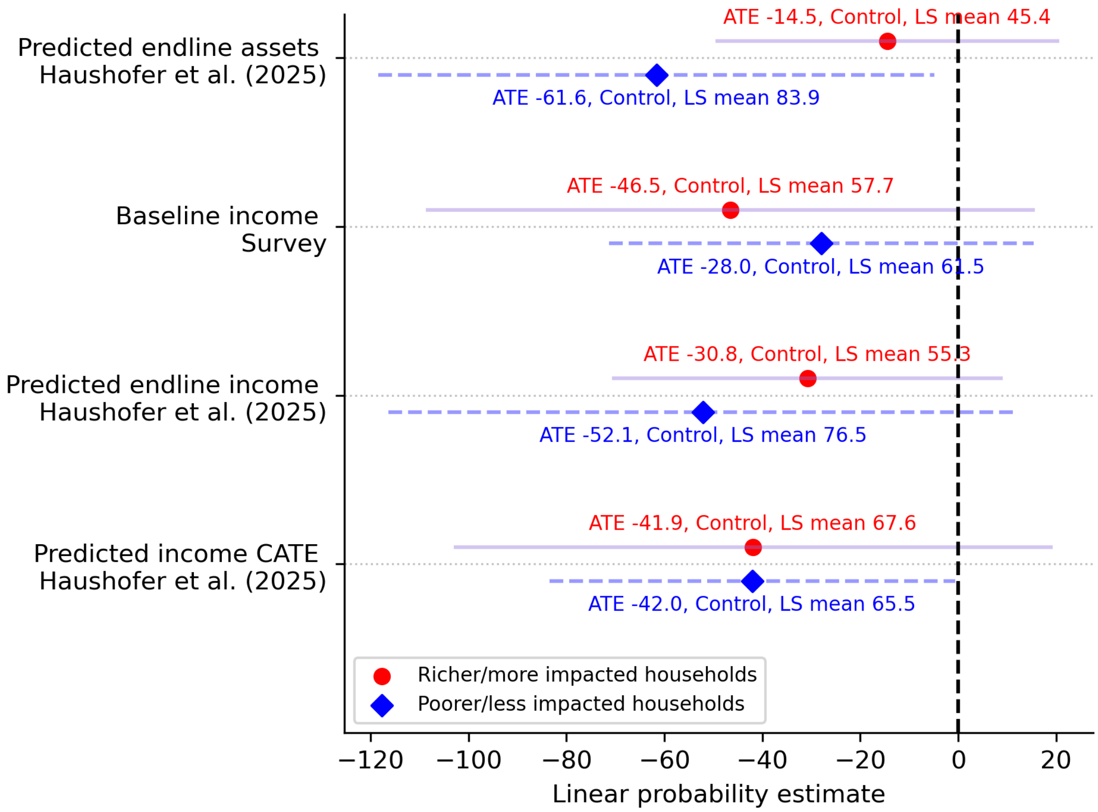
Panel A: Infant Mortality Means By Month



Panel B: Birth Month Fixed Effects

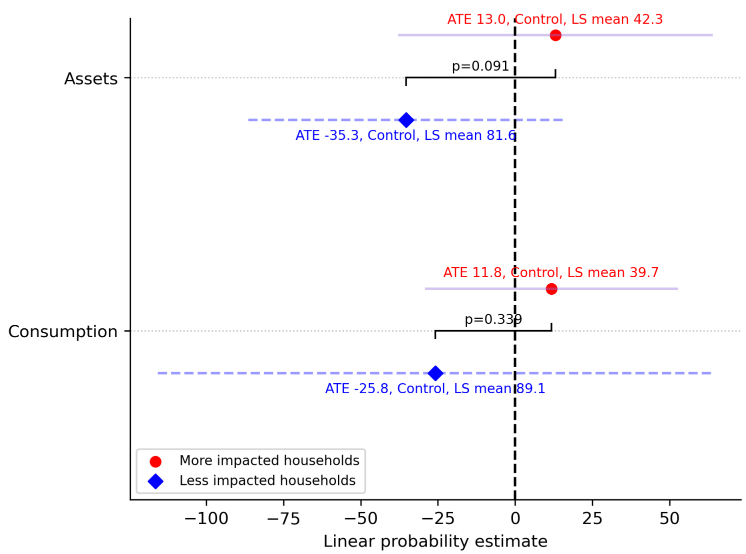
Notes: This figure is based on the main EL3 birth census sample. Panel A reports the mean infant mortality rate by month across 2011-14, 2015-17, and 2018-19. Panel B reports birth month fixed effects controlling for birth year fixed effects, with September as the omitted month. The lean season is defined as April-August using the definition from Burke, Bergquist, and Edward Miguel (2019), whose study area comprised a nearby region in rural Kenya. The years 2020 and 2021 are omitted as COVID-19 pandemic impacted the region, though patterns appear similar if those years are included. 95% confidence intervals are reported in Panel B.

Figure C.7: Heterogeneous effects by deprivation and impact: additional measures

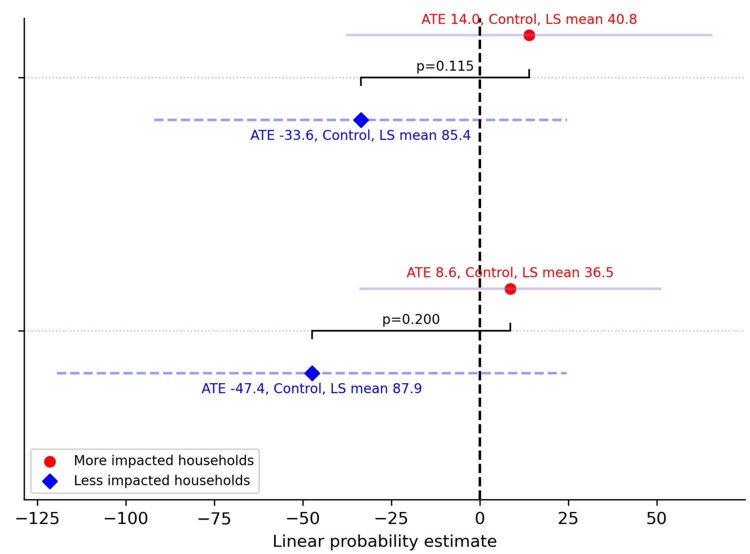


Notes: Transfer-eligible households visited in the baseline survey and EL3 census are included. The plotted effects are treatment effects on infant mortality (per 1000) births, based on the indicated heterogeneity cuts. Predicted endline assets and income use per-capita generalized random forest (GRF) predictions from Haushofer et al. (2025), where "poorer" households are defined as those where the share of time the household is classified as "most deprived" exceeds the median. Baseline income refers to baseline household income measured by the survey, where poorer households are those where this value falls below the median. The predicted income CATE row classifies households as "more impacted" if the share of time they were classified as "most impacted" in terms of per capita income CATEs in Haushofer et al. (2025) exceeds the median. Stars denote the significance of the difference between richer/poorer or less/more impacted households. Figure 3.4 reports effects in terms of consumption and assets, where there is more predictable heterogeneity.

Figure C.8: Heterogeneous effects by deprivation and impact: controlling for left out heterogeneity dimension



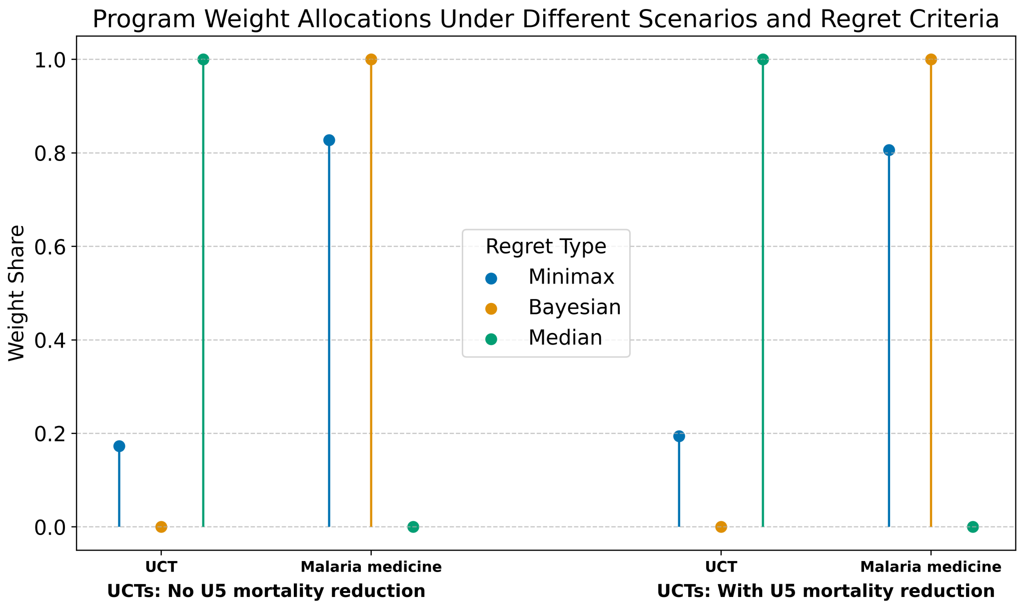
Panel A: Heterogeneity by baseline socioeconomic status



Panel B: Heterogeneity by predicted conditional average treatment effect (CATE)

Notes: This figure estimates the same regressions as Figure 3.4, but estimates of heterogeneity by deprivation control for impacted and impacted interacted with treatment and vice-versa. Transfer-eligible households visited in the baseline survey and EL3 census are included. Panel A reports effects on infant mortality (per 1000 births) by below vs above-median wealth, estimated using Eq. (3.1). In Panel A, richer households by the measure of "assets" reported above median baseline household assets, while "consumption" uses the predicted endline per capita generalized random forest (GRF) prediction from Haushofer et al. (2025). Consumption is predicted since it was not measured at baseline. A household is defined as "richer" if the share of GRF runs where it was defined as "most deprived" per the Haushofer et al. (2025) was lower than average. Panel B reports effects on infant mortality (per 1000) births based on whether the predicted conditional average treatment effect (CATE) of per capita assets or consumption was lower or higher. Specifically, a household is defined as "more impacted" if the share of GRF splits in Haushofer et al. (2025) that classified it as "most impacted" exceeds the sample median, across the full sample of births among eligible HHs. Panel A controls for the corresponding measures (and their interactions with treatment) in Panel B, and Panel B controls for the corresponding measures and treatment interactions in Panel A.

Figure C.9: Share of Funds Allocated to UCTs and Malaria Medicine by Regret Type



Notes: This figure reports the share of funds allocated to UCTs or Malaria medicine by an agent under minimax, Bayesian (mean), and median regret excluding and including the mortality reduction benefits of UCTs. The agent faces uncertainty about the value of a statistical life (VSL), and they minimize the given regret function over the posterior distribution of VSL. We treat other parameters as known. We use GiveWell's estimate of the cost per life saved from Malaria medicine to estimate its returns. The posterior distribution of VSL estimates in this sample obtained from Killeen (2025), Kremer, Leino, et al. (2011), Berry, Fischer, and Guiteras (2020), and Redfern et al. (2019) using Bayesian hierarchical meta analysis with a log-normal prior.

Table C.1: Descriptive Statistics for the Census of Births

	(1) Infant Mortality	(2) Child Mortality	(3) Neonatal Mortality	(4) Maternal Age at Birth
Births to All Households (2011-23)				
Mean	33.71	46.69	15.11	25.65
Observations	101,336	101,348	100,498	95,353
Births to Households Present at Baseline (2011-23)				
Mean	33.61	46.40	18.50	25.72
Observations	77,847	77,849	77,229	73,247
Births to Eligible Baseline Households, UCT Period (2015-17)				
Mean	34.54	50.68	19.23	25.55
Observations	6,334	6,335	6,240	6,003
Births to Eligible Baseline Households, Pre-UCT Period (2011-14)				
Mean	35.43	54.18	14.11	24.73
Observations	8,634	8,635	8,504	8,208
Births to Eligible Baseline Households, Post-UCT Period (2018-21)				
Mean	37.89	51.54	20.09	26.35
Observations	9,577	9,577	9,457	9,129
Births to Eligible Baseline Households (2011-21)				
Mean	36.15	52.25	17.77	25.57
Observations	24,545	24,547	24,201	23,340
Births to Ineligible Baseline Households (2011-21)				
Mean	31.28	45.44	17.39	25.60
Observations	40,872	40,872	40,712	38,205

Notes: This table reports means for descriptive characteristics and observation counts for groups in the EL3 census of births. The four descriptive characteristics reported are the infant under-one mortality rate (in deaths per thousand births), the child under-five mortality rate (deaths per thousand births), the neonatal mortality rate (deaths per thousand births), and the mean maternal age at birth. The five groups in the census for which these data are presented are as follows: all births recorded in the census (across 2021-23), all births to households present at baseline (2011-23), births to transfer-eligible households present at baseline across 2015-17 (our main sample under study), births to eligible households present at baseline across 2011-14, births to eligible households present at baseline across 2018-21, births to eligible baseline households across 2011-21, and births to ineligible baseline households across 2011-21.

Table C.2: Balance on Tracking, Study Area Presence, and Survey Completion

	(1) Household Tracked (Census)	(2) Present in Study Area (Census)	(3) Completed Census (Census)	(4) Household Tracked (Survey)	(5) Completed Survey (Survey)
Treatment Village	0.001	-0.006	0.006	0.005	0.008
	(0.006)	(0.004)	(0.004)	(0.005)	(0.007)
High-Saturation Sublocation	-0.006	0.008*	-0.005	-0.003	0.004
	(0.008)	(0.004)	(0.006)	(0.004)	(0.007)
Treatment Villages in High-Saturation Sublocations	-0.005	0.002	0.001	0.002	0.012
	(0.009)	(0.004)	(0.007)	(0.005)	(0.009)
Control Mean	0.897	0.915	0.905	0.965	0.893
Observations	62742	52582	48125	8372	8075

Notes: This table reports the share of households present at baseline (2014-15) tracked in the third endline (EL3) census, present in the study area at EL3, completed the EL3 census, tracked in the EL3 survey, and completed the EL3 survey, as well as balance by treatment status. The share tracked in the census (column 1) is reported as a share of baseline households non-deceased in the second endline census (EL3). The share present in study area (column 2) is reported as a share of non-deceased baseline households tracked in EL3. The share completing the EL3 census (column 3) is reported as a share of baseline households present in the study area, excluding holiday homes which do not comprise a household's primary residence. The share tracked in the EL3 survey (column 4) is reported as a share of baseline households sampled for the EL3 survey. The share completing the EL3 survey (column 5) is reported as a share of baseline households tracked in the survey. Observations encompass split households, which were created since 2014-15 from the 65,383 parent households originally present at baseline. Standard errors are clustered at the sublocation level. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table C.3: Transfers and Mortality Including Secondary Outcomes, 2015-17

	Primary outcomes		Secondary outcomes	
	(1) Infant Mortality 2015-17	(2) Under-5 Mortality 2015-17	(3) Neonatal Mortality 2015-17	(4) Days survived under-5 2015-17
Panel A: Reduced-form results				
Treatment village	-5.74	-11.96*	-3.56	17.63*
	(5.85)	(6.38)	(4.00)	(10.22)
MHT adjusted p-value	[0.234]	[0.110]		
High-saturation sublocation	-12.13**	-5.68	-6.00	11.78
	(5.04)	(6.66)	(3.75)	(10.50)
Treatment village in high-saturation sublocation	-17.87***	-17.64***	-9.56**	29.41***
	(4.94)	(5.86)	(3.80)	(9.75)
Panel B: Spatial IV results				
Spatial IV: Own village effect	-7.98*	-12.72**	-3.56	17.83*
	(4.82)	(5.55)	(3.62)	(9.55)
Spillover effect	-11.49*	-12.91	-11.00*	26.94**
	(6.84)	(8.12)	(6.16)	(13.12)
ATE on eligibles	-19.46***	-25.63***	-14.57**	44.78***
	(6.94)	(8.54)	(5.75)	(13.47)
MHT adjusted p-value	[0.044]	[0.036]		
Control Mean	40.21	57.37	23.05	1,735.17
Observations	6,317	6,318	6,237	6,257

Notes: This table is based on the main EL3 birth census sample and includes estimated effects on pre-specified secondary outcomes. Reduced form standard errors are clustered at the sublocation level. Spatial HAC standard errors (Conley, 2008) with a cutoff of 10km are reported for IV estimates. MHT corrected p-values in brackets. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table C.4: Unconditional Cash Transfers and Infant Mortality, 2011-21

	(1) 2011-14	(2) 2015-16	(3) 2017 (Drought)	(4) 2018-21
Panel A: Reduced-Form Estimates				
Own village	0.03	-3.61	-9.84	-4.11
	(4.66)	(7.87)	(9.01)	(4.42)
High-saturation spillovers	5.08	-11.44	-13.50	0.72
	(4.79)	(7.18)	(8.19)	(4.44)
ATE in high-saturation sublocations	5.11	-15.04***	-23.35**	-3.39
	(6.31)	(5.69)	(10.39)	(4.60)
Control mean	34.53	32.97	54.87	41.92
Observations	8605	4216	2101	9544
Panel B: Spatial IV Estimates				
Own village	1.09	-6.28	-13.08	-4.28
	(4.68)	(6.26)	(8.63)	(4.35)
High-saturation spillovers	-5.22	-6.38	-17.05	4.85
	(5.01)	(7.97)	(10.67)	(5.71)
ATE in high-saturation sublocations	-4.13	-12.66	-30.13***	0.57
	(6.42)	(8.34)	(11.59)	(6.82)
Control mean	34.53	32.97	54.87	41.92
Observations	8605	4215	2101	9544

Notes: Effects are estimated using Equation (3.1) (Panel A) and (3.2) (Panel B) on transfer-eligible women from the main EL3 census sample. Infant mortality estimated effects are reported per 1,000 live births. The periods examined are 2011-14 in column 1 (the period before UCTs were distributed), 2015-16 (the first two years of the UCT period), 2017 (the final year of the UCT period, during which there was a major drought in Kenya), and 2018-21 (the period following UCT disbursal). The main results for the pre-specified 2015-17 period of central interest are reported in Table 3.1. The ATE in high-intensity villages equals the average total effect of own-village estimates and spillovers in high-saturation sublocations. Reduced form standard errors are clustered at the sublocation level. Spatial HAC standard errors with a cutoff of 10km are reported for IV estimates. * $p < .10$, ** $p < .05$,

Table C.5: Transfers and Mortality Excluding Stillbirths, 2015-17

	Reduced-form		Spatial IV	
	(1) Infant Mortality 2015-17	(2) Under-5 Mortality 2015-17	(3) Infant Mortality 2015-17	(4) Under-5 Mortality 2015-17
Own village	-6.85	-13.10**	-9.02**	-13.79***
	(5.53)	(6.13)	(4.44)	(5.26)
High-saturation spillovers	-9.63**	-3.18	-8.35	-9.84
	(4.84)	(6.52)	(5.47)	(7.33)
ATE in high-saturation sublocations	-16.48***	-16.28***	-17.38***	-23.63***
	(4.64)	(5.60)	(5.73)	(8.11)
Percent reduction in HS sublocations	52.04%	33.24%	54.88%	48.24%
Control Mean	31.66	48.98	31.66	48.98
Observations	6,266	6,267	6,266	6,267

Notes: This table is based on the main EL3 birth census sample and estimates effects on infant and child mortality excluding stillbirths. The sample excludes 51 stillbirths that were classified as such by the SmartVA verbal autopsy algorithm as opposed to directly reported by the household as a stillbirth. Reduced form standard errors are clustered at the sublocation level. Spatial HAC standard errors (Conley, 2008) with a cutoff of 10km are reported for IV estimates. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table C.6: Mortality Effects of Cash Transfer Spillovers on Non-Recipients

	(1) Infant Mortality 2015-17	(2) Under-5 Mortality 2015-17	(3) Neonatal Mortality 2015-17	(4) Days survived under-5 2015-17
Total effect IV	0.74	2.20	-4.04	-1.15
	(3.84)	(4.96)	(3.01)	(7.27)
Control Mean	31.82	46.84	19.21	1,753.45
Observations	13,191	13,192	13,102	13,154

Notes: This table reports estimates of spillover effects of the cash transfers on non-recipients of cash (both ineligible households and eligible households in control villages) on mortality for births across 2015-2017. Estimates are constructed using a spatial instrumental variable approach similar to Equation 3.2. Spatial HAC standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table C.7: Spillover Effects of Transfers for Eligibles and Poor Ineligibles

	Infant Mortality		Hospital Delivery	
	(1) Eligibles	(2) Poor Inelig.	(3) Eligibles	(4) Poor Inelig.
Total effect IV	-19.46***	-8.51*	0.20**	0.23**
	(6.50)	(5.09)	(0.07)	(0.09)
Direct	-7.98*		0.01	
	(5.21)		(0.03)	
Spillovers	-11.49*	-8.51*	0.18***	0.23**
	(5.98)	(5.09)	(0.07)	(0.09)
Control Mean	40.21	32.23	0.44	0.53
Observations	6317	7150	1154	290

Notes: Effects are estimated using Equation (3.2) augmented with an interaction with the characteristic of interest. Poor ineligibles are defined as ineligible households (i.e., home does not have a thatched roof) whose home has both mud walls and a mud floor. Information on home attributes is sourced from the baseline census, and approximately 67% of ineligible households fulfill these criteria. The median eligible household in the control group possessed USD PPP 435 in net non-land, non-housing assets, whereas the median poor ineligible household in the control group held USD PPP 643 in assets (48% higher than eligibles). In contrast, the median control household among other ineligibles possessed USD PPP 1335 in assets (108% higher than poor ineligibles and 207% higher than eligibles). All outcome variables are for births and households across 2015-17. Spatial HAC standard errors (Conley, 2008) with a cutoff of 10km are reported for IV estimates. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table C.8: Heterogeneity by Gender, Maternal Age, and Birth Order

	Infant Mortality			**Child Mortality**		
	(1) Male	(2) Maternal Age 25+	(3) Birth Order 2+	(4) Male	(5) Maternal Age 25+	(6) Birth Order 2+
Treatment Village	-12.80*	-5.80	-14.34*	-12.92*	-12.92*	-22.77***
	(7.02)	(7.26)	(8.39)	(7.09)	(7.09)	(8.27)
High-Saturation Sublocation	-6.11	-7.42	-2.34	1.35	1.35	2.74
	(6.98)	(6.06)	(7.62)	(7.56)	(7.56)	(9.32)
Treatment Village x Heterogeneity	8.53	-4.57	9.98	-3.06	-3.06	13.63
	(9.04)	(9.57)	(9.26)	(10.32)	(10.32)	(10.54)
High Saturation Sublocation x Heterogeneity	-8.45	-5.90	-13.68	-10.36	-10.36	-11.34
	(9.09)	(9.71)	(10.64)	(11.00)	(11.00)	(11.51)
Treatment Villages in High-Saturation Sublocations	-18.91***	-13.22*	-16.68**	-11.56	-11.56	-20.03**
	(6.75)	(7.44)	(7.52)	(8.52)	(8.52)	(8.96)
Treatment in High-Saturation Sublocations Including Heterogeneity Characteristic	-18.83***	-23.69***	-20.39***	-24.99***	-24.99***	-17.74**
	(7.01)	(7.29)	(6.70)	(8.48)	(8.48)	(7.62)
Control Mean	39.95	39.95	39.95	57.00	57.00	57.00
Observations	6256	6331	6256	6332	6332	6257

Notes: This table is based on the main EL3 birth census sample. Effects are estimated using Equation (3.1) augmented with an interaction with the characteristic of interest. Maternal age at birth is analysed using pre-specified five-year age groups (under 20, 20-24, 25-29, 30-34, and 35 and above); this analysis splits on the median five-year group, with 46% of births having a mother below the age of 25. Child birth order is calculated using the order of births by mother across the 2011-23 period and does not take into account births to mothers prior to 2011 as information on those births was not obtained in the census. Standard errors clustered at the sublocation level. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table C.9: Unclassified Verbal Autopsies by Treatment Status

	(1) Undetermined VA (2011-14)	(2) Undetermined VA (2015-17)	(3) Undetermined VA (2018-21)
Treatment Village	0.06	-0.10*	-0.03
	(0.05)	(0.05)	(0.05)
High-Saturation Sublocation	0.03	0.08	0.03
	(0.05)	(0.06)	(0.05)
Treatment Villages in High-Saturation Sublocations	0.09 (0.06)	-0.02 (0.07)	0.00 (0.05)
Control Mean	0.11	0.21	0.16
Observations	263	201	325

Notes: This table reports regressions exhibiting the association between treatment status and the probability that a completed verbal autopsy (VA) had an undetermined cause by the SmartVA classification algorithm for transfer-eligible under-5 deaths across three periods: 2011-14 (pre-UCT), 2015-17 (the period contemporaneous with UCT disbursal), and 2018-21 (post-UCT). Standard errors are clustered at the sublocation level. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table C.10: Unconditional Cash Transfers and Fertility

	(1) Births Per 100 Women 2011-14	(2) Births Per 100 Women 2015-17	(3) Births Per 100 Women 2018-23	(4) Births Per 100 Women 2015-23	(5) Completed Fertility Women 45+ in 2023 2011-23
Panel A: Probability of Birth					
Total Effect IV	0.28	0.79**	-0.00	0.13	-0.08
	(0.25)	(0.34)	(0.18)	(0.13)	(0.07)
Control Mean	7.19	7.54	7.01	5.58	7.69
Observations	23597	23597	23597	23597	979
Panel B: Number of Births					
Total Effect IV	0.22	1.13***	-0.15	0.27	0.28
	(0.34)	(0.38)	(0.31)	(0.28)	(0.78)
Control Mean	9.12	8.48	9.66	9.27	11.31
Observations	23597	23597	23597	23597	979

Notes: Effects are estimated using Equation (3.2) on transfer-eligible women from the main EL3 census sample. Fertility outcomes represent the annual probability a woman gives birth to at least one child (Panel A), and the number of births per 100 women in a given period, with birth rates standardized to a one-year window to facilitate comparisons (Panel B). Column 5 restricts the sample to transfer-eligible women who were aged 45 year or over in 2023, the year in which the EL3 census was conducted. Spatial HAC standard errors in parentheses. * $p < .10$, ** $p < .05$,

Table C.11: Cross-Sectional Relationship Between Socio-Demographic Status and Infant Mortality in Control Eligible Households

	(1) Under-1 Mortality	(2) Under-1 Mortality	(3) Under-1 Mortality	(4) Under-1 Mortality
Transfer-Eligible	10.54**	10.58**		
	(5.25)	(5.31)		
Below-Median Baseline Assets			50.88**	60.51**
			(23.74)	(25.34)
Mean in Low SES Group	39.95	39.95	81.38	81.38
Mean in High SES Group	29.40	29.40	30.50	30.50
Observations	5202	5202	489	489
Birth Characteristic Controls	NO	YES	NO	YES

Notes: This table reports regressions exhibiting the cross-sectional association between infant mortality and markers of socio-demographic status for births to eligible households in control low-saturation villages across 2015-17. Columns 1-2 present the association between eligibility for the unconditional cash transfers (determined by having a thatched roof) and infant mortality. Columns 3-4 present the association between holding a level of assets at baseline below the median level and infant mortality within the sample of transfer-eligible households surveyed in the representative baseline survey sample. Columns 1 and 3 present bivariate relationships, whereas Columns 2 and 4 include pre-specified controls for birth and maternal characteristics (i.e., gender, year of birth, and maternal age). Robust standard errors are utilized. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table C.12: UCTs and Household Labor Supply by Gender Around the Time of Birth

	Spatial IV		**Reduced-Form**	
	(1) Women Hours Worked Last Week	(2) Men Hours Worked Last Week	(3) Women Hours Worked Last Week	(4) Men Hours Worked Last Week
Panel A: First 6 Months In-Utero				
Treated High-Saturation Estimate	6.57	-18.93	-0.79	-0.03
	(16.96)	(16.05)	(6.13)	(7.00)
Control Mean in Period	38.44	46.23	38.44	46.23
Panel B: 3rd Trimester and First 3 Months Postpartum				
Treated High-Saturation Estimate	-24.48**	-2.83	-15.06***	0.53
	(12.18)	(9.84)	(5.29)	(5.00)
Control Mean in Period	39.51	39.30	39.51	39.30
Panel C: Next 6 Months Postpartum				
Treated High-Saturation Estimate	-5.08	1.88	9.98	-5.44
	(13.21)	(8.93)	(8.20)	(5.11)
Control Mean in Period	53.26	26.49	53.26	26.49
Control Mean Overall	45.85	34.77	45.85	34.77
Total Observations	1766	1677	1766	1677

Notes: This table encompasses households surveyed in the first endline survey (2016-17) with a recorded birth in the EL3 census. Labor supply encompasses hours worked across the household (in agricultural employment, non-agricultural self-employment, and wage employment), as well as hours spent searching for work, in the week prior to the first endline survey. The table displays the estimated differential effect of the cash transfers on labor supply for three groups: households with a woman in the first 6 months of pregnancy when surveyed, households with a woman in the third trimester of pregnancy or who gave birth in the past 3 months when surveyed, and households with a woman who gave birth 3-9 months ago when surveyed. Treatment effects are estimated separately for each group using Equation 2 (columns 1-2) and Equation 1 (columns 3-4) augmented with interactions with the group of interest. Spatial HAC standard errors (columns 1-2) and robust standard errors clustered at the sublocation level (columns 3-4) in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table C.13: UCTs and Effects on Maternal Health and Nutrition

	(1) Health Index Female All	(2) Health Index Female Child $\leq$ 1	(3) Health Index Male All	(4) Health Index Male Child $\leq$ Age 1	(5) Child Food Security All	(6) Adult Food Security All
Total Effect IV	0.06	0.28**	0.00	-0.08	0.17***	0.09*
	(0.06)	(0.11)	(0.07)	(0.12)	(0.06)	(0.05)
Control Mean	-0.02	0.04	0.14	0.25	-0.10	0.03
Observations	3234	730	1521	299	3673	4768

Notes: * $p < .10$, ** $p < .05$, *** $p < .01$. The health-status index is a weighted, standardized average of self-reported health, index of health symptoms, and experienced a major health problem since baseline. The food security index is a weighted, standardized average of food security outcomes including number of days in the past week meals were skipped, gone entire days without food, and gone to bed hungry. The construction of both indices was pre-specified. The sample is restricted to eligible households.

Table C.14: UCTs and Effects on the Health of Women in Households with Infants

	(1) Health Index Female Child ≤ 1	(2) Major Health Episode Female Child ≤ 1	(3) Recent Health Symptoms Child ≤ 1	(4) Current Self-Reported Health Female Child ≤ 1	(5) Health Episode Resolved Female Child ≤ 1
Total Effect IV	0.28**	-0.09**	-0.22**	0.11	-0.10
	(0.11)	(0.04)	(0.11)	(0.15)	(0.15)
Control Mean	0.04	0.14	0.00	3.65	0.32
Observations	730	730	730	730	85

Notes: * $p < .10$, ** $p < .05$, *** $p < .01$. The health-status index is a weighted, standardized average of self-reported health, index of health symptoms, and experienced a major health problem since baseline. The food security index is a weighted, standardized average of food security outcomes including number of days in the past week meals were skipped, gone entire days without food, and gone to bed hungry. The construction of both indices was pre-specified. The sample is restricted to eligible households.

Table C.15: Association Between Treatment Status and Baseline Characteristics For Eligible Households with 2015-17 Births

	(1) Education	(2) Age	(3) Marital Status	(4) Income	(5) Assets	(6) Household Size
Total Effect IV	0.05	-5.08	0.06	-1649.02	-7425.11	0.09
	(0.33)	(4.18)	(0.04)	(2452.02)	(4574.36)	(0.28)
Control Mean	8.50	26.71	0.72	12342.35	33897.07	4.35
Observations	1353	1353	1353	1353	1353	1353

Notes: This table encompasses transfer-eligible households which reported a birth from 2015-17 in the EL3 census and were surveyed at baseline. Regressions are reported exhibiting the association between treatment status and six baseline household socio-demographic characteristics (maximum years of education of household members, average household age, number of adults, number of children, baseline household assets, and baseline household income). Spatial HAC standard errors (Conley, 2008) with a cutoff of 10km are reported for IV estimates. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table C.16: Unconditional Cash Transfers and Actual versus Predicted Fertility

	Actual Fertility Among Adult Women			Predicted Fertility Among Mothers	
	(1)	(2)	(3)	(4)	(5)
	2011-14	2015-17	2018-21	2015-17	2018-21
Total Effect IV	0.28	0.79**	0.02	0.70	-0.15
	(0.25)	(0.34)	(0.27)	(0.86)	(0.60)
Control Mean	7.19	7.54	8.31	8.86	9.16
Observations	23597	23597	23597	1353	1353

Notes: Columns 1-3 are estimated on transfer-eligible women from the EL3 census. Columns 4-5 are estimated on transfer-eligible women from the EL3 census from households surveyed at baseline. Fertility outcomes represent the annual probability a woman gives birth to at least one child. The last two columns report predicted fertility among women actually giving birth in a given period, based on a random forest model trained on baseline survey data from women in control, low-saturation villages with six household socio-demographic characteristics used as predictors: baseline household income, baseline household assets, maximum years of education of household members, average household member age, marital status of the primary respondent, and household size. The random forest is trained using five-fold cross-validation, and the model is estimated on indicators capturing all combinations of the six socio-demographic variables, which are each binned into quartiles. Spatial HAC standard errors in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table C.17: Association Between Treatment Status and Predicted Probability of Giving Birth Among Mothers Giving Birth Over 2015-17

	(1) Probability of Birth Among Actual Mothers Socioeconomic Variables	(2) Probability of Birth Among Actual Mothers LASSO Variables
Total Effect IV	0.70	0.11
	(0.86)	(0.10)
Control Mean	8.86	7.11
Observations	1353	1353

Notes: This table encompasses mothers in transfer-eligible households present at baseline sampled in the baseline survey, and reports regressions exhibiting the association between treatment status and the probability of birth among mothers who actually gave birth over 2015-17. In column 1, the probability of birth is predicted using six baseline household socio-demographic characteristics (maximum years of education of household members, average household member age, marital status of primary respondent, baseline household income, baseline household assets, and household size) using a random forest trained on women in control, low-saturation villages. In column 2, the probability of birth is predicted linearly using LASSO from a set of 243 baseline household- and village-level characteristics. Spatial HAC standard errors (Conley, 2008) with a cutoff of 10km are reported for IV estimates. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table C.18: Impacts on Infant Mortality in Household Survey Sample

	(1) Pre-Specified Controls	(2) Socioeconomic Controls	(3) PDS LASSO-Selected Controls
Panel A: Spatial IV Estimates			
Total Effect IV	-36.16*	-35.41*	-37.83**
	(18.81)	(19.16)	(18.70)
Panel B: Reduced-Form Estimates			
Treatment Village	-25.76*	-26.13*	-26.48**
	(12.97)	(13.24)	(12.94)
High-Saturation Sublocation	-10.97	-8.10	-11.96
	(11.81)	(11.83)	(11.60)
Treatment Villages in High-Saturation Sublocations	-36.73** (17.91)	-34.23* (17.90)	-38.44 (17.95)
Control Mean	65.83	65.83	65.83
Observations	1486	1486	1486

Notes: This table is based on births reported in the EL3 birth census to transfer-eligible households present at baseline and surveyed in the baseline household survey. Effects estimated by Equation (3.2) (Panel A) and Equation (3.1) (Panel B) of the cash transfer on infant mortality rates are reported. Mortality data are sourced from the endline 3 census whereas household- and village-level control variables are sourced from the baseline survey. Column 1 presents infant mortality results in the baseline survey sample with only pre-specified (e.g., year of birth fixed effects, birth gender, mother age group) covariates. Column 2 additionally includes controls for six baseline household socioeconomic characteristics (maximum years of education of household members, primary respondent age, marital status of primary respondent, income, assets, and household size), along with the pre-specified controls. Column 3 includes controls selected by PDS LASSO, with a total of 243 baseline variables available for selection, along with pre-specified controls. For consistency across the two equations, we use the PDS LASSO covariates selected for Equation (3.1) for Equation (3.2) as well. Spatial HAC standard errors (Conley, 2008) with a cutoff of 10km are reported for IV estimates, and standard errors clustered at the sublocation level are reported for reduced-form estimates.* $p < .10$, ** $p < .05$, *** $p < .01$.

C.2 Comparison with Non-Experimental Variation

The existence of birth census data over the period 2011-2023 enables us to benchmark the experimental cash transfer treatment effect estimates against several dimensions of non-experimental variation in economic circumstances. In short, across multiple sources of variation, child survival appears to respond substantially to economic conditions in rural Kenya, with non-experimental differences exhibiting the same sign and similar magnitudes as the main cash transfer experimental results reported above.

Cross-Sectional Variation

One form of variation is cross-sectional: transfer-eligible households were notably poorer on average than ineligible households (as noted above), as measured by either consumption or assets. As illustrated by Columns 1-2 of Appendix Table C.11, the birth census indicates that infants born to ineligible mothers in control villages were substantially less likely to die in their first year of life than children from eligible households in these same villages, but that in the 2015-17 period of transfer disbursal, the treatment effect from transfers is sufficient to close the infant mortality gap between treated eligible and ineligible households. Across 2015-17, under-one mortality among control eligible births was 10.5 deaths per thousand higher than among control ineligibles, a 36% difference statistically significant at the five percent level. This socioeconomic gradient in mortality also holds even when controlling for birth characteristics such as maternal age and gender, and a similar pattern is evident when examining under-one mortality or all the years across the birth census as opposed to the main cash transfer period. Furthermore, even within eligible households, births to households with below-median assets at baseline are substantially less likely to survive their first year in control villages: Columns 3-4 of Appendix Table C.11 show that infant mortality among households with below-median assets is 2.7 times higher (81.4 deaths per thousand births) than among households with above-median assets (30.5 deaths per thousand). Of course, there are many other interpretations of this pattern and many potential omitted variables or confounders — which is precisely why the current study exploits experimental variation to estimate causal impacts — but we note that these cross-sectional observational patterns remain broadly consistent with the large experimental estimates described above.

Seasonality in Infant Mortality

Child mortality displays a similarly high degree of sensitivity to inter-temporal changes in economic conditions. The information on exact month of birth for most children in the birth census allows us to estimate month of birth effects for both infant and child mortality (controlling for several other characteristics such as birth year, gender, and maternal age). Burke, Bergquist, and Edward Miguel (2019) define the pre-harvest "lean season" in rural western Kenya as encompassing the months of April through August, after which the main grain harvest occurs. The analysis of the birth census during the pre-COVID period of

2011-19 reveals a meaningful degree of seasonality in infant mortality: rates of mortality for infants born in August (often the very peak of the lean season just before the harvest) are 21.9 deaths per thousand births higher than for infants born in the next month (Appendix Figure C.6). This difference – which is similar in magnitude to the cash transfer treatment effect we estimate – is statistically significant at the one percent level, and forms part of a larger pattern of mortality rising as the lean season progresses: from May through August, infant mortality rates rise by an average of 6.44 deaths per thousand each month. Under-five survival rates display virtually identical patterns.

Aggregate Economic Shocks

Mortality in the birth census also responds to major aggregate economic shocks: as is evident in Panel A of Figure 3.1, infant mortality rises across the board in 2017, a year in which a major drought affected rural Kenya (BBC, 2017). Furthermore, the COVID-19 pandemic struck Kenya in early 2020, leading to a large recession as well as reduced access to many public services. The birth census data indicates that infant mortality nearly doubled from 2019 to 2020, with the largest jump occurring precisely in April 2020, the first full month of pandemic related lockdowns. Mortality remains elevated in 2021, a year in which the COVID-19 pandemic continued to affect Kenya and another severe drought impacted the region (World Bank, 2022).

As data were collected on exact date of birth for the majority of children censused, we can more precisely examine the impact of the COVID-19 shock on child health outcomes. On March 27, 2020, the Kenyan government imposed a strict lockdown intended to combat the spread of COVID-19, including a nationwide curfew during evening and night hours (Achuka and Gisesa, 2020). Individuals who left their homes during the curfew were subject to a maximum of three months imprisonment, a Ksh. 1000 (USD PPP 23.5) fine, or both, and local media reported numerous instances of police brutality meted out on Kenyans breaking curfew (Mireri, 2020).

A regression discontinuity analysis indicates that the infant mortality rate sharply increased immediately after the lockdown. Panel B of Figure C.5 illustrates that infant mortality doubled in the study region the week after the lockdown, a result statistically significant at the five percent level. These effects are likely caused by changes to health care access rather than COVID-19 deaths per se: reported COVID-19 deaths exhibit no such discontinuity at the lockdown date, and mortality from COVID-19 is largely concentrated among the elderly as opposed to children (Ma et al., 2021). Furthermore, the jump in infant mortality is particularly high in areas near a physician-staffed health facility, suggesting that a decline in access to care may have contributed to part of the deterioration in child health outcomes. The sharp rise in infant mortality in the study setting during the COVID-19 lockdown is consistent with non-experimental evidence from India suggesting that mortality among infants at six months rose 44% during India's lockdown, with the authors positing that reduced healthcare access and utilization may have driven this decline in child survival (Asker, Dhongde, and Shonchoy, 2025).

Cross-Country Association Between Income and Health

We last examine the cross-country relationship between per capita GDP and infant mortality. As previously documented by others, income and health are closely associated with each other (Preston, 1975; Cutler, Lleras-Muney, and Vogl, 2012). Data on per capita GDP is obtained from the World Bank, 2024, whereas infant mortality estimates are sourced from the United Nations, 2024.[1] In 2014, the year in which cash transfers in our study region commenced, each log point increase in GDP per capita was associated with a 0.79 log point reduction in infant mortality, a relationship statistically significant at the one percent level and with an R-squared of 0.74. As seen in Panel A of Figure C.5, augmenting the study region's GDP by the same proportion as the cash transfers did for treated households would result in a 36% decline in infant mortality as predicted by the global cross-country GDP-mortality relationship. While the cross-country association is correlational rather than causal in nature, the magnitude of this predicted reduction is notably similar to our experimental estimates.

In summary, child mortality appears to be highly sensitive to economic conditions in this study's context, irrespective of the source of variation. Non-experimental approaches recover similar magnitudes as the large experimental effects estimated in this study.

C.3 Details of Benefit-Cost Calculations

This appendix details the estimates presented in Section 3.7. We first describe how the estimated posterior distribution of VSLs presented in Figure 3.8 is constructed and detail the benefit curves in the figure. We then describe the econometric decision theory problem producing the weights plotted in Figure C.9.

The posterior distribution of VSLs is constructed using Bayesian hierarchical meta analysis. We begin by positing that VSLs in countries with similar income levels follow a log-normal distribution with weakly informative priors. The log-normal distribution imposes that VSL is never negative and captures the possibility of a large right tail. We assume that the location parameter of the distribution is normally distributed with a mean of 10 and standard deviation of 10, and the scale parameter is 5 times a half-Cauchy distribution with location parameter 0 and scale parameter 0.5. A half-Cauchy distribution with a scale parameter of 0 and scale of 0.5 is a recommended best practice for imposing weakly informative priors ("A Practical Guide to Random-Effects Bayesian Meta-Analyses With Application to the Psychological Trauma and Suicide Literature" 2022). We multiply these draws by 5 to ensure priors do not favor VSL values near 0. Estimates of the posteriors are not sensitive to these decisions, consistent with the fact that the prior is only weakly informative.

We then use Markov Chain Monte Carlo methods in the Python package "pymc" to estimate the posterior distribution. We do so by matching the likelihood to estimates from

[1] We also utilize an estimate of the study region's GDP, which is obtained from the Kenya National Bureau of Statistics, 2019.

Killeen (2025), Kremer, Leino, et al. (2011), Berry, Fischer, and Guiteras (2020), and Redfern et al. (2019). We utilize these estimates because they were obtained from populations with similar income levels. For instance, we exclude León and Edward Miguel (2017b) because the population studied has a much higher income level, and Killeen (2025) shows that this predicts a VSL orders of magnitude higher. We also restricted our focus to revealed preference estimates, other than Redfern et al. (2019), since they reflect real world choices rather than hypothetical questions that may be prone to social desirability bias.

Figure 3.8 then plots welfare gains of UCTs and other programs by VSL. For UCTs, we plot estimates without child mortality reductions priced in imposing the 2.5 consumption multiplier from Egger, Haushofer, Edward Miguel, Niehaus, et al. (2022b), the same with our estimates of the cost per life saved of the UCTs, and targeted transfers to pregnant women that have a consumption multiplier of 1 with and without lives saved priced in. We also plot the benefit curve by VSL of GiveWell's top recommendation, the Malaria Consortium, since this is one of the most effective health interventions and GiveWell reports estimates of the cost per life saved. We assume no direct consumption benefit and therefore all benefits are (1,000/cost per life saved) x VSL.

Figure C.9 plots the share of funds that a decision maker would invest in UCTs or the Malaria Consortium by regret type in the following econometric decision problem.

A social planner allocates a budget across J programs, each characterized by its consumption gain and expected lives saved per dollar spent. The planner's objective is to minimize regret given uncertainty about the true VSL, which we will denote by $\theta \in \Theta$.

Each program $j \in \{1, \ldots, J\}$ has two attributes: g_j, the consumption gain per dollar spent, and S_j, the expected number of lives saved per dollar spent.

The planner chooses a weight vector $\mathbf{w} = (w_1, \ldots, w_J)$ such that $\sum_{j=1}^{J} w_j = 1$, where w_j represents the share of the budget allocated to program j.

The true VSL, denoted θ, is uncertain and follows a distribution $\theta \sim F(\theta)$. We use the estimated posterior distribution of θ in calculations, but in minimax regret impose that it falls within the 95% confidence interval since one would otherwise mechanically invest only in the most effective health program since VSL has no upper bound in the log-normal prior assumed. For simplicity, and since we were unable to find confidence intervals on the cost per life saved of the Malaria Consortium program, we take point estimates of consumption multiplier and cost per life saved as given.

For a given realization of θ, the optimal allocation $\mathbf{w}^*(\theta)$ maximizes total benefit:

$$\mathbf{w}^*(\theta, \mathbf{S}) = \arg\max_{\mathbf{w}} \sum_{j=1}^{J} w_j \left(S_j \theta + g_j\right).$$

The regret for a given $(\theta, \mathbf{S})$ and choice $\mathbf{w}$ is:

$$R(\mathbf{w}, \theta) = \max_{\mathbf{w}'} \sum_{j=1}^{J} w'_j \left(S_j \theta + g_j\right) - \sum_{j=1}^{J} w_j \left(S_j \theta + g_j\right).$$

We consider three regret criterion. First, the minimax regret criterion selects $\mathbf{w}$ to minimize the worst-case regret:

$$\min_{\mathbf{w}} \sup_{\theta} R(\mathbf{w}, \theta).$$

where we use the 5th and 95th percentile of VSL values estimated in the posterior distribution to define the respective worst case draws of θ for each program. This case reflects the optimal allocation for an extremely risk averse decision maker that wishes to maximize welfare in the worst possible case.

Second, the Bayesian (average) regret criterion minimizes the expected regret given the posterior distribution of VSL:

$$\min_{\mathbf{w}} \mathbb{E}_{\theta} \left[R(\mathbf{w}, \theta) \right].$$

In practice, we use 1,000 posterior draws from the posterior to estimate this value. This value would be optimal for a risk neutral decision maker that wishes to maximize expected welfare gains, but is heavily influenced by the outlying right tail of VSL values. In practice, this results in the Malaria Consortium's program receiving 100% of investment in all cases, even though it is dominated by UCTs for 75% of VSL values simulated from the posterior distribution.

Third, we consider the allocation that minimizes the median regret over θ:

$$\min_{\mathbf{w}} \text{median}_{\theta} R(\mathbf{w}, \theta).$$

This provides a regret measure that is less sensitive to outliers in VSL.

We construct these estimates including and excluding the benefits of mortality reductions from UCTs. In practice, this has little effect on results since the cost per life saved is lower under the Malaria Consortium than UCTs.