Too often governments fail to provide access to quality public services to the poor. My work focuses on this issue and the bottlenecks that impede high-quality government provision of education, health care, and environmental protection. Within education, I have projects studying public-private partnerships in fragile states, teacher incentive payments, and extra educational resources. Within health care, I am mainly interested in public health insurance plans, including how they affect pharmaceutical prices. Within environmental protection, I am interested in mining and regulation, including how mining affects public health and how regulations affect the prevalence of illegal mining. In conjunction with this empirical research agenda, I work on methodological issues in applied econometrics and statistics.

Education In my job market paper, *"Outsourcing Service Delivery in a Fragile State: Experimental Evidence from Liberia"*, Justin Sandefur (Center for Global Development), Wayne Sandholtz (UC San Diego), and I study whether outsourcing public services to private entities improves service delivery in fragile states. We provide experimental evidence from the Partnership Schools for Liberia (PSL) program, which delegated management of 93 public schools to eight different private organizations. In total, schools in our sample account for 3.4% of all public primary schools, serving 8.6% of students enrolled in public early childhood programs and primary. We randomize treatment at the school level using a match-paired design to study heterogeneity across service providers. We sample students from pre-treatment enrollment records, which allows us to provide clean estimates of the program's intention-to-treat (ITT) effect on learning, uncontaminated by endogenous student selection. Within one academic year, outsourcing increased students scores in English and math by $.18\sigma$, relative to control schools. Our design allows us to study heterogeneity across providers: While the highest-performing providers generated increases in learning of 0.27σ, the lowest-performing providers had no impact on learning. Consistent with the rules of provider contracts, we find no evidence that providers engaged in student selection. However, providers were allowed to shift pupils from oversubscribed schools and underperforming teachers to other government schools. These results suggest that leveraging the private sector to improve service delivery in fragile states is promising, but they also highlight the importance of procurement rules and contracting details to aligning public and private interests.

Three more experiments are underway, with the field work component finished. In *"Inputs, Incentives, and Complementarities in Primary Education: Experimental Evidence from Tanzania"*, Karthik Muralidharan (UC San Diego), Isaac Mbiti (University of Virginia), and I study complementarities across policies in education. While the idea that complementarities across policies can lead to increasing returns has a long tradition in economics, there is limited evidence that clearly identifies such complementarities. We present evidence of the impact of providing schools with (a) unconditional capitation grants, (b) bonus payments to teachers based on student performance, and (c) both of the above. We find no impact on student learning from providing either the grants or teacher incentives but significant positive effects from providing both. We find strong evidence of complementarities between improving school inputs and teacher incentives, with the combined effect being greater than the sum of the individual effects.

In designing teacher incentive schemes, policymakers often face trade-offs between ease-of-implementation and theoretical grounding (efficacy). In the project described above, we chose an incentive design for teachers that was "implementable" at scale. The bonuses paid to teachers were based on the number of children who exceeded an absolute threshold of learning, as opposed to improvements in learning. However, this could induce teachers to neglect students who are well above or well below the passing threshold. In a follow-up project ("*Designing Teacher Performance Pay Programs: Experimental Evidence from Tanzania*") we compare the pay-for-percentile scheme proposed by Barlevy and Neal (2012) to a simple threshold design. Despite the theoretical advantages of the pay-for-percentile system — it induces optimal effort among

teachers — the threshold system (easier to implement at scale) is at least as effective in raising student learning in this setting.

My work in Tanzania highlights the importance of tailoring teaching to the student's learning level. In *"Cross-Age Tutoring: Experimental Evidence from Kenya"* I partner with a chain of private schools to study cross-age tutoring — in which older students tutor younger students — as an inexpensive alternative for providing personalized instruction. I find that tutoring in math has a small positive effect on math test scores. The effect is concentrated among middle-ability students, suggesting that tutors are not able to help advanced learners and those lagging behind grade-level competencies.

Health and Environment My non-experimental work has focused on other issues affecting public services in developing countries. I have two projects on Colombia's public health insurance plan. My most complete paper (*"Benefit Plans, Insurer Competition, and Pharmaceutical Prices: Evidence from Colombia"*) studies how the public plan's coverage choices affect drug prices. I propose a simple theoretical framework that illustrates that listing a drug might increase or decrease its price, operating through two distinct mechanisms. Using a difference-in-difference strategy, I find empirically that listing decreases prices for most drugs but consistent with theory, it increases them for drugs that face no competition.

Another area of my research is how environmental policies affect health outcomes. Santiago Saavedra (Universidad del Rosario) and I study how gold mining affects public health and how regulations affect the prevalence of illegal mining. In *"The Effect of Gold Mining on the Health of Newborns"* we develop a novel instrument for newborns' exposure to mining impacts. We find that mining, and especially illegal mining, negatively affects newborns. In *"Local Incentives and National Tax Evasion: The Response of Illegal Mining to a Tax Reform in Colombia"* we develop a comprehensive dataset using satellite imagery and novel machine-learning techniques to measure the prevalence of illegal mining. We then show that a recent municipal funding reform increased illegal mining, which, given our earlier work, has important implications for public health.

Methods In conjunction with my empirical research, I have worked on misapplication of instrumental variables; approaches to replication in economics; and cross-cut designs and their implications in field experiments.

A common assumption implicit in most papers using instrumental variables (IV) to assess the effect of air pollution on health is that a single pollutant is endogenous. However, pollutants are often co-produced, implying that IV estimates will be biased. In *"Using Instrumental Variables under Partial Observability of Endogenous Variables for Assessing Effects of Air Pollution on Health"*, Prashant Bharadwaj (UC San Diego), Tarik Benmarhnia (UC San Diego), and I document this issue in the literature, and show that IV estimates may be more biased than OLS estimates.

Replication is a critical component of scientific credibility, as it increases our confidence in the reliability of knowledge generated by original research. In "Improving The Effectiveness of Replication in Economics", Paul Gertler (UC Berkley), Sebastian Galiani (University of Maryland), and I examine replication rates in economics and outline a new model for journals to take over the task of replication after acceptance and before publication.

Finally, Karthik Muralidharan, Kaspar Wuthrich (UC San Diego), and I are developing best practices for implementing cross-cut designs (commonly used in field and lab experiments). We document the prevalence of two-step model selection in these studies, characterize the inference problems that result, and compare a number of possible approaches.