Long-Term Impact of Development Research: Evidence from Kenya

Vaidehi Bulusu [1]

April 26, 2024

University of California, Berkeley
Department of Economics

Spring 2024 Undergraduate Senior Honors Thesis
Advised by Professor Marco Gonzalez-Navarro

Abstract

Randomized controlled trials are extensively used in development economics to identify the causal impact of an intervention on outcomes of interest, usually health and education. RCTs are usually focused on the short-term impact of the intervention, hence a natural question to ask is whether this development research activity positively impacts development outcomes over the long-term. In this paper, I use a differences-in-differences model to study the impact of research activity (RCTs in particular) in Kenyan provinces on health- and education-related outcomes over the 2003 to 2022 time period. I find statistically significant effects on some education and health outcomes.

Keywords: RCTs, development interventions, development outcomes, development research, health, education, long-term, long-run

[1](Email: vaidehi-b@berkeley.edu) I would like to thank Professor Marco Gonzalez-Navarro for his generous support. All errors are my own.

1 Introduction

Economic development is a highly active area of research, concerned with answering questions about how to improve various outcomes in low- and middle-income countries. Of particular interest in economic development research is conducting randomized controlled trials (RCTs) to estimate the causal effect of various interventions – such as the free distribution of a health product or the implementation of a cash transfer policy – on outcomes related to development, particularly health and education (often referred to as 'human capital'). RCTs are particularly effective for estimating the causal effect of one variable on another, as they involve the random assignment of the treatment across the treatment and control groups, thereby eliminating confounding variables. They also allow researchers and policymakers to develop solutions tailored to the region of interest. Hence, RCTs revolutionized the field of development economics, with Abhijit Banerjee, Esther Duflo, and Michael Kremer winning the 2019 Nobel Prize for their implementation of RCTs to alleviate poverty.

Many of these studies found that the intervention they were studying significantly improved outcomes of interest. For instance, Miguel and Kremer (2004) conducted an RCT in Kenyan schools to evaluate the impact of a deworming drug on schoolchildren. They found that the treatment significantly improved health and school participation for both treated and untreated schoolchildren, concluding that these benefits (including externalities, which were previously underestimated), justify full subsidization of the treatment. Another important study is Cohen and Dupas (2010) which investigated whether cost-sharing of insecticide-treated bednets (ITNs) was necessary to avoid wastage. They found that cost-sharing does not reduce wastage, nor does it result in selection of women who need the product more. On the contrary, it actually reduces demand, depriving it from women who need it. Hence, their findings justify the free distribution of ITNs.

While a particular RCT is useful for estimating causal impact, it is usually conducted within a relatively short timeframe, so it the benefits observed in the study are over the short- or medium-term. Hence, a natural question to ask is whether these benefits 'stick':

that is, whether or not the improvements in these outcomes persist over the long-term. This is the question I attempt to answer in my paper. Much of development research has been conducted in Africa and South Asia, with a significant portion of this research taking place in Kenya. Hence, I will be focusing on Kenya for my analysis. In particular, my research question is: what is the long-term effect of RCTs on health and education outcomes in Kenya?

While there have been impact evaluations conducted by organizations such as the World Bank for their development projects, few studies have analyzed the long-term impact of RCTs. One study that does this is Bouguen et al. (2019), which analyzed RCTs to identify which ones are amenable to long-term follow-up. Follow-up studies would allow researchers to generate high-quality evidence on the long-term effects of development research. In this paper, I attempt to fill a gap in existing literature by looking at the collective effect of RCTs as opposed to the effect of specific RCTs. Hence, this is essentially a question about whether high research activity in a region has a positive effect on development outcomes. For example, a significant portion of RCTs conducted in Kenya were conducted in the Western province. I aim to answer the question of whether this region has better health and education outcomes compared to other provinces, holding other factors constant.

Additionally, this is also a question about how research translates into the real world. There are two causal mechanisms through which research can have a long-term impact at a larger scale. First, the study induces behavioral change, which could result in a significant effect if the sample is large enough (especially if the study has externalities, like the worms study). The second is that the study induces policy change. For example, both Miguel and Kremer (2004) and Cohen and Dupas (2010) concluded that free subsidization of their respective products is justified based on their results; however, was this actually implemented? Since my paper is a province-level analysis and RCTs are usually conducted on a smaller scale, the second mechanism is a more plausible explanation for any significant effects.

To answer my question, I used data on RCTs from the American Economic Association (AEA) RCT Registry and data on health and education variables across Kenyan provinces from the Demographic and Health Surveys (DHS) Spatial Data Repository. I analyzed projects that took place in the 2003 to 2022 time period using a differences-in-differences model. I ran two specifications: one that looked at the effect of the total number of projects, and another which looked at the effect of the average coverage (a measure of how long projects ran for on average). I conducted this data at the province-level due to a lack of data at the county-level.

The rest of this paper proceeds as follows. In Section 2, I will provide an overview of the most relevant literature on this topic, particularly, papers estimating the long-term effects of RCTs in development research, as well as papers on evidence-based policymaking. In Section 3, I will describe the data and its limitations, and provide a descriptive analysis of the variables. In Section 4, I will discuss my estimation strategy and present results from regression models. Finally, I will conclude with a summary and discussion of key results, and propose directions for further research.

2 Existing Evidence on the Impact of RCTs

A significant number of RCTs were conducted in the late 90s and early 2000s, so enough time has passed for us to think about the long-term impact of this research. Bouguen et al. (2019) exploits this fact to analyze RCTs from development economics, focusing on cash transfers and child health programs. They first surveyed RCTs that already conducted follow-ups and found that a significant portion of them have positive long-term impacts. For example, Hoddinott et al. (2008) studied the impact of improving nutrition in early childhood and found that it positively affected outcomes, such as wages, in adulthood. They also identified studies that are amenable to follow-up research. More generally, they provide a framework for assessing the long-term impact of studies that were already conducted, and ways of designing future studies that would make long-term follow-up feasible, including collecting panel data, improving participant tracking, and access to

additional data (e.g. cell phone usage). They argue that focusing on follow-ups would provide researchers an opportunity to generate high-quality evidence on the benefits of RCTs.

Along the same lines is Greenhalgh et al. (2016), which aims to precisely define the term 'research impact' and reviews approaches for measuring it. The authors conclude that most accurate assessment methodologies are labor-intensive, making them infeasible to conduct. This could explain the lack of long-term follow up discussed in Bouguen et al. (2019).

However, researchers have in recent years shifted their attention to the long-run benefits generated by their studies. An example is Chowdhury (2021) which evaluates a large-scale anti-poverty intervention conducted in Bangladesh. They find that the study had a significant impact on asset accumulation, but that there was heterogeneity in the distribution of these benefits on households.

These papers relate to how individual RCTs can be evaluated for long-term impact. A paper that looks at the collective effect of development research is Kremer et al. (2018). In particular, they analyze the impact of a broader category of studies (referred to as 'development innovations'), including RCTs as well as development aids provided by organizations such as the United Nations and World Bank. They develop a method for measuring portfolio return for development innovation portfolios, and estimated that the Development Innovation Venture's (DIV)[2] portfolio returned $5 per dollar spent and a social rate of return of 77%. They conclude that their results justify increasing spending on international development.

Apart from effort on the part of researchers to perform long-term follow-ups and evaluations, another mechanism through which RCTs have a long-term impact is if they are translated into policy. For instance, Cohen and Dupas (2010) concluded that their results justify the free distribution of ITNs based on the lives they save. Had the government implemented a program in which it handed out ITNs to pregnant women for free, we would

[2]DIV is a program of the United States Agency for International Development that funds development ventures: https://www.usaid.gov/div.

surely observe the positive effects found in the study on a larger scale (holding other factors constant). Hence, a lack of statistically significant results could be explained by a lack of evidence-based policymaking. Some prominent studies that discuss this include DellaVigna et al. (2022), Baron (2018), Caves and Lueling (2021) and Vivalt and Coville (2023), all of which discuss the limitations to evidence-based policymaking.

My paper attempts to fill a gap in the existing literature by modelling the relationship between research activity and development outcomes.

3 Data and Descriptive Analysis

3.1 Data Background and Preparation

This analysis requires province-level panel data on RCTs and development outcomes during a significant time period. I found data on RCTs on the AEA RCT Registry for RCTs conducted since 1998. This website is the American Economic Association's registry for randomized controlled trials. It was established in April 2012 but allows institutions to register their trials from earlier as well. It collects data on 50 different variables, including start and end dates for trials, start and end dates for interventions, keywords relevant to the study, experimental design, clustering, and number of observations.

I obtained data on outcome variables from the DHS Spatial Data Repository. This repository provides nationally representative data on various variables related to development including nutrition, mortality rates, fertility rates, child health, gender equality, and educational outcomes for various years. The data is geographically-linked at the national as well as sub-national level so it can be mapped in a geographic information system (GIS). For Kenya, the Repository has yearly data across the 1989 to 2022 time period (at the province-level until 2014 and at the county-level from 2014 onwards, which is a limitation – this is discussed further below).

To prepare the dataset for analysis, I first downloaded data on all RCTs (a total

of 8327 trials) and filtered for trials related to Kenya (250 trials remained)[3]. After performing text analysis to extract province information and filtering out projects without this information, 136 trials remained across 8 provinces. I also collected province-level data on variables related to health and education (described below) for 1998, 2003 and 2022.

I constructed two treatment variables: number of projects and coverage. The number of projects is the total number of projects that took place in each province and is a measure of the level of research activity in a province. The coverage for a project is the portion of the 2003-2023 time period (so 20 years) for which the project ran, and is a measure of the level of exposure the region had to projects (as it is reasonable to expect that projects which ran for a longer period of time have more sustained impact). This variable is then averaged across provinces to obtain average coverage.

Since I am using a differences-in-differences model, I have discretized the treatment variables into 3 categories: low intensity, medium intensity, and high intensity. If a province has a total of 5 projects or fewer, or an average coverage of 5% or lower, it is categorized as low-intensity (this is the 'control' group). If a project has more than 5 projects but fewer than 20 projects, or an average coverage of more than 5% but fewer than 10%, it is categorized as medium-intensity. All other provinces (i.e. provinces with more than 20 projects or an average coverage of more than 10%) are categorized as high-intensity.

3.2 Data Limitations

Here I describe the constraints on data affecting the analysis that follows. The first is that the RCT data is not comprehensive as not all studies are registered. Institutions are required to register only if they are submitting their paper to AEA. Moreover, the Registry doesn't have standardized guidelines on encoding location information: some projects provided information at the city-level, others at the county-level, and most at the

[3]I filtered based on the 'Country names', 'Title,' and 'Abstract' columns.

province-level. Many projects did not have information location at all or had extremely general descriptions (e.g. 'Africa'). As a result, projects whose province I could not ascertain were filtered out. These two factors could potentially result in selection bias of the sample of projects being analyzed.

An additional constraint is a lack of county-level data. Ideally, this analysis would be performed at the county-level for more precise estimates. However, Kenya established counties in 2010 , so county-level data is not available before then. Moreover, county-level data on DHS is available only starting from 2014, but 2014-2022 is not a long enough time period for this analysis.

3.3 Variable Description

The following is a description of the outcome variables that will be used for this analysis. I chose health and education outcomes, and the subset of these outcomes described below, as they have been the focus of most RCTs.

Table 1: Description of Outcome Variables

Education	
EDEDUCWCPR	Percentage of women with completed primary education
EDEDUCWCSC	Percentage of women with completed secondary education
EDEDUCMCPR	Percentage of men with completed primary education
EDEDUCMCSC	Percentage of men with completed secondary education
Health	
CHVACCCBAS	Percentage of children (12-23 months) who received all 8 basic vaccinations
CMECMRCCMR	Child mortality rate
CNNUTSCHA2	Children stunted[4]
CNNUTSCWA2	Children underweight[5]

3.4 Descriptive Analysis

First, I provide some descriptive statistics on the treatment variable. Below is a table showing the distribution of the number of projects and average coverage across the provinces. Note that a project can take place in multiple provinces hence the number of

projects doesn't sum to 136.

Table 2: Distribution of the Treatment Variables across the Provinces

Province	**Number Projects**	**Coverage**
Central	10	14.43%
Coast	2	0%
Eastern	14	7.62%
Nairobi	52	5.82%
Northern	5	25%
Nyanza	21	9.75%
Rift Valley	4	8.75%
Western	44	12.19%

The most projects were conducted in the Nairobi Province, followed by the Western Province. The Nairobi county is responsible for almost all the projects in the province, as the pre-filtered dataset contained 56 projects conducted in the Nairobi province. Projects in the Western Province take place across more counties, with Busia (38 projects), Kakamega (29 projects), and Bungoma (28 projects) being the counties with the most research activity. The Northern and Central provinces had the highest coverage, at 25% and 14.43% respectively. There is more variation in the number of projects than coverage, indicating that provinces differ significantly in the level research activity taking place but less so in the amount of time for which those studies occur on average.

Below are summary statistics on pre- and post-treatment education data.

Table 3: Pre- and Post-Treatment Means for Education Variables

Variable	**count**	**mean**	**std**	**min**	**25%**	**50%**	**75%**	**max**
Pre-Treatment (2003) Data								
EDEDUCMCPR	8	21.537	7.576	8.8	18.75	20.6	24.65	32.9
EDEDUCMCSC	8	15.412	5.749	7.2	12.5	14.5	17.95	26.3
EDEDUCWCPR	8	22.425	9.69	1.8	21.1	22.9	27.1	33.6
EDEDUCWCSC	8	11.4	7.759	0.3	7.575	10.15	13.05	26.1
Post-Treatment (2022) Data								
EDEDUCMCPR	8.0	17.0531	4.654	9.7	14.128	18.693	20.185	22.56
EDEDUCMCSC	8.0	20.073	5.318	14.85	17.356	18.043	20.936	30.9
EDEDUCWCPR	8.0	18.129	5.981	5.3	16.242	20.175	21.804	23.333
EDEDUCWCSC	8.0	17.527	6.591	7.566	14.906	16.562	19.370	28.6

The table shows that the average percentage of men who completed primary school and average percentage of women who completed primary school decreased during this

time period. On the other hand, the average percentage of men who completed secondary school and average percentage of women who completed secondary school has increased during this time period. In both pre- and post-treatment data, there is a disparity along gender lines as a higher percentage of men have completed primary and secondary education than women. However, this gender disparity has reduced since 2003. Estimating the effect of development research on gender disparity in various outcomes is a potential avenue for further research. Overall, these statistics show a favorable trend in the variables.

Below are summary statistics on pre- and post-treatment health data.

Table 4: Pre- and Post-Treatment Means for Health Variables

Variable	**count**	**mean**	**std**	**min**	**25%**	**50%**	**75%**	**max**
Pre-Treatment (2003) Data								
CHVACCCBAS	8.0	48.025	19.954	7.5	42.225	53.2	60.	70.8
CMECMRCCMR	8.0	45.000	28.804	10.0	26	35.5	72.25	84.0
CNNUTSCHA2	8.0	33.912	5.971	23.5	30.3	35.7	37.55	41.2
CNNUTSCWA2	8.0	16.137	7.1257	5.4	11.675	16.6	18.85	29.6
Post-Treatment (2022) Data								
CHVACCCBAS	8.0	76.146	17.872	33.533	76.935	80.478	84.775	89.175
CMECMRCCMR	8.0	8.784	5.011	4	5.943	6.952	9.575	17.75
CNNUTSCHA2	8.0	16.365	3.997	11.1	13.82	14.879	20.141	22.08
CNNUTSCWA2	8.0	10.361	5.033	5.12	5.425	10.056	14.495	17.833

The table shows that on average, the percentage of children who received all 8 basic vaccines has increased, and child mortality rate, percentage of children stunted, and percentage of children underweight has reduced. Overall, the trend in these variables is favorable.

4 Methodology and Results

4.1 Empirical Approach

To estimate the causal effect of research activity on education and health outcomes, I used a differences-in-differences model. I found evidence of parallel trends based on a

formal test, which justifies this model (see Appendix A). I ran two specifications varying the treatment variable, and in each specification, ran four panels varying the outcome variable.

I am estimating the following model:

$$Y_{it} = \alpha + \beta_1 Treat_{1i} + \beta_2 Treat_{2i} + \beta_3 Post_t + \beta_4 Treat_{1i} \cdot Post_t + \beta_5 Treat_{2i} \cdot Post_t + \epsilon_i, \quad (1)$$

where Y_{it} is the health- or education-related outcome variable, $Treat_{1i}$ is the dummy variable indicating units categorized as medium-intensity, $Treat_2 i$ is the dummy variable indicating units categorized as high-intensity, and $Post_t$ is the dummy variable indicating whether the observation is in 2022 (the pre-treatment period is 2003). The low-intensity group serves as the control/reference group and is omitted.

The parameters I am interested in estimating are β_4 and β_5, which are the differences-in-differences estimators for the medium- and high-intensity groups respectively. For evidence against the null, we would expect to see positive and statistically significant estimates for the the percentage of women and men who completed primary and secondary school, and the percentage of children who received all 8 basic vaccines; and negative and statistically significant estimates for the child mortality rate, percentage of children who are stunted, and percentage of children who are underweight.

A limitation of this analysis is the small sample size. The 136 RCT studies are aggregated across provinces resulting in 8 units. This could lead to imprecise estimates. Moreover, I am only looking at four variables related to health and education each, so we cannot draw conclusions on the relationship between research activity and education/health in general based on these results.

4.2 Results

Tables 5 and 6 present the results from the differences-in-differences regression for the education and health variables, respectively.

Table 5: Differences-in-Differences Regression Results for Education

Panel A: EDEDUCMCPR

	(1)	(2)
Intercept	26.6***	30.5***
	(3.9)	(0.0)
Medium	-0.45	-9.57***
	(7.80)	(1.11)
High	-6	-3.8
	(3.9)	(6.2)
Post	-7.73	-7.94***
	(5.37)	(0.0)
Medium x Post	1.44	5.14**
	(8.65)	(2.06)
High x Post	6.28	0.62
	(5.45)	(6.31)
N	12	12
R^2	0.453	0.644
Adjusted R^2	-0.002	0.347
Residual Standard Error	42.62 (df = 6)	120.75 (df = 6)

Panel B: EDEDUCMCSC

	(1)	(2)
Intercept	16.65***	17.6***
	(0.95)	(0.0)
Medium	-0.7	-3.63***
	(3.2)	(1.01)
High	-4.35***	-2.45
	(1.38)	(3.85)
Post	2.08*	0.32***
	(1.25)	(0.0)
Medium x Post	3.38	4.17***
	(4.98)	(1.2)
High x Post	3.13	4.53
	(2.3)	(6.43)
N	12	12
R^2	0.617	0.402
Adjusted R^2	0.297	-0.096
Residual Standard Error	94.88 (df = 6)	746.65 (df = 6)

Panel C: EDEDUCWCPR

	(1)	(2)
Intercept	22.9***	21.9***
	(1)	(0.0)
Medium	9.7***	3.83
	(1.41)	(3.47)
High	-2.4	4.55
	(1.56)	(7.15)
Post	-4.32**	-1.5***
	(2.08)	(0.0)
Medium x Post	-5.98**	-3.77
	(2.51)	(4.13)
High x Post	5.46*	-3.34
	(2.94)	(7.34)
N	12	12
R^2	0.919	0.295
Adjusted R^2	0.852	-0.293
Residual Standard Error	238.13 (df = 6)	854.59 (df = 6)

Panel D: EDEDUCWCSC

	(1)	(2)
Intercept	10.6***	9.9***
	(0.7)	(0.0)
Medium	3.75	-0.27
	(4.01)	(1.44)
High	-3.15***	3.1
	(0.74)	(5.3)
Post	4.61*	3***
	(2.41)	(0.0)
Medium x Post	1.6	4.25***
	(6.37)	(1.5)
High x Post	4.21*	4.26
	(2.52)	(7.07)
N	12	12
R^2	0.717	0.621
Adjusted R^2	0.481	0.305
Residual Standard Error	544.15 (df = 6)	2470.22 (df = 6)

Note: The dependent variable in Panel A is the percentage of men who completed primary education. The dependent variable in Panel B is the percentage of men who completed secondary education. The dependent variable in Panel C is the percentage of women who completed primary education. The dependent variable in Panel D is the percentage of women who completed secondary education. *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively. The estimation is performed using OLS regression.

Table 6: Differences-in-Differences Results for Health

Panel A: CHVACCCBAS

	(1)	(2)
Intercept	56.8***	62.1***
	(5.3)	(0.0)
Medium	8.25	-14.27
	(7.82)	(9.12)
High	-17.75**	-4
	(8.27)	(12.7)
Post	21.23***	17.82***
	(5.63)	(0.0)
Medium x Post	-1.41	14.56
	(8.91)	(9.45)
High x Post	26.04***	13.02
	(8.95)	(12.7)
N	12	12
R^2	0.930	0.801
Adjusted R^2	0.871	0.635
Residual Standard Error	42.62 (df = 6)	120.75 (df = 6)

Panel B: CMECMRCCMR

	(1)	(2)
Intercept	29**	41***
	(12)	(0.0)
Medium	-9.5	2.33
	(15.31)	(23.82)
High	48***	-1
	(13.89)	(30)
Post	-21.41*	-33.4***
	(12)	(0.0)
Medium x Post	7.68	-0.2
	(15.31)	(24.05)
High x Post	-38.96***	4.98
	(13.94)	(30.63)
N	12	12
R^2	0.925	0.406
Adjusted R^2	0.862	-0.088
Residual Standard Error	94.88 (df = 6)	746.65 (df = 6)

Panel C: CNNUTSCHA2

	(1)	(2)
Intercept	39.1***	41.2***
	(2.1)	(0.0)
Medium	-3.95	-3.8***
	(4.56)	(1.091)
High	-3.4	-7.95***
	(2.12)	(2.15)
Post	-17.72***	-19.12***
	(2.21)	(0.0)
Medium x Post	-0.76	0.012
	(5.67)	(2.6)
High x Post	-3.1	0.32
	(2.32)	(2.4)
N	12	12
R^2	0.946	0.965
Adjusted R^2	0.901	0.937
Residual Standard Error	10.88 (df = 6)	6.98 (df = 6)

Panel D: CNNUTSCWA2

	(1)	(2)
Intercept	18.9***	19***
	(0.1)	(0.0)
Medium	-4.6	-3.03
	(3)	(2.46)
High	-5.05**	-5.4**
	(2.05)	(2.3)
Post	-4.37***	-4.54***
	(0.12)	(0)
Medium x Post	-1.25	-0.659
	(4.66)	(4)
High x Post	-2.81	-2.56
	(2.38)	(2.68)
N	12	12
R^2	0.782	0.655
Adjusted R^2	0.601	0.368
Residual Standard Error	9.11 (df = 6)	14.42 (df = 6)

Note: The dependent variable in Panel A is the percentage of children who received all 8 vaccinations. The dependent variable in Panel B is the child mortality rate. The dependent variable in Panel C is the percentage of children who are stunted. The dependent variable in Panel D is the percentage of children who are overweight. *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively. The estimation is performed using OLS regression.

The results show that being in the medium-intensity coverage group was associated with a significantly higher percentage of men completing primary school (5.14% higher, significant at the 10% level) and secondary school (4.17% higher, significant at the 1% level), and with a significantly higher percentage of women completing secondary school (4.25% higher, significant at the 1% level). Additionally, being in a high-intensity projects group was associated with a significantly higher percentage of women completing primary school (5.46% higher, significant at the 10% level) and secondary school (4.21% higher, significant at the 10% level). It is surprising that we observe some these results for the medium as opposed to high-intensity treatment groups. Moreover, the significance is not consistent across specifications – for panels A, B and D, the coverage variable had statistically significant estimates, while for panel C, the number of projects variable had statistically significant estimates. Another surprising result is that being in a medium-intensity projects group is associated with a significantly lower percentage of women completing primary school (5.98% lower, significant at the 5% level). The small sample size could explain these observations.

For the health variables, being in a high-intensity projects group was associated with a significantly higher number of children receiving all 8 basic vaccines (26.04% higher, significant at the 1% level), and with a significantly lower child mortality rate (38.96% lower, significant at the 1% level). No other statistically significant estimates were observed.

Based on these preliminary results, there is some evidence that RCTs have a long-term impact on some education and health outcomes.

5 Discussion and Conclusion

Considered the gold standard for causal identification, RCTs allow us to estimate the effect of a treatment on an outcome. Development economics makes extensive use of RCTs, with numerous studies being conducted primarily in Africa and Asia. The amount of research done in these regions has only grown in the past few decades, which raises

the question of the impact this research activity had at a larger scale in the long-run. With meta-data available on these studies and outcome variables, I attempt to answer this question.

Using a differences-in-differences estimator, I find evidence that research activity has a significant long-term impact on some education and health variables. In particular, I find that the coverage of projects has a positive effect on primary and secondary school completion rate for men, and secondary school completion rate for women. I also found that the number of projects has a positive effect on the primary and secondary school completion rate for women, and vaccination rate for children, as well as a negative effect on child mortality rate. Some of the results were surprising; in particular, the medium-intensity group having a more significant impact than the high-intensity group on some education variables. Note that this is preliminary evidence limited by the small sample size of 8 provinces.

Building on this paper, future research could focus on replicating this analysis at a more precise-level based on the availability of data, potentially at the county- or even city-level. It would also be interesting to heterogeneity in the long-term benefits, as Chowdhury (2021) did in their paper, across genders for instance. To investigate the indirect effects of RCTs, we can look at the impact of specific categories of projects on seemingly unrelated outcomes (for example, the effect of projects focusing on workplace gender equality on health outcomes), using keywords to categorize projects. Additionally, to analyze the extent to which RCTs induce behavioral change, we can look at the effect of the share of population treated (using data on sample size from the RCT database) on long-term outcomes.

There are two main causal mechanisms underlying the relationship between research activity and development outcomes. The first is the extent to which RCTs induce behavioral change. As discussed by Bouguen et al. (2019), designing RCTs in a way that makes long-term follow-up feasible and following through with the follow-up would contribute to sustained behavioral change by participants in the study. If the study is conducted

on a large scale and has externalities, these effects can be detected. A more plausible mechanism for RCTs to have a large-scale and long-term effect is through the transfer of the results from studies to policy. This underlies the importance of eliminating barriers to evidence adoption, as DellaVigna et al. (2022) discuss in their paper. These factors are incredibly important to ensure that research in development economics contributes to meaningful and sustained improvements in crucial outcomes.

References

AEA RCT Registry (2012). https://www.socialscienceregistry.org/trials/search. Online; accessed 26 April 2024.

Banerjee, A. V. and Duflo, E. (2007). The economic lives of the poor. *Journal of Economic Perspectives*, 21(1):141–167.

Baron, J. (2018). A brief history of evidence-based policy. *The ANNALS of the American Academy of Political and Social Science*, 678(1):40–50.

Bertrand, M., Duflo, E., and Mullainathan, S. (2002). How much should we trust differences-in-differences estimates? Working Paper 8841, National Bureau of Economic Research.

Bouguen, A., Huang, Y., Kremer, M., and Miguel, E. (2019). Using randomized controlled trials to estimate long-run impacts in development economics. *Annual Review of Economics*, 11(1):523–561.

Callaway, B., Goodman-Bacon, A., and Sant'Anna, P. H. C. (2024). Difference-in-differences with a continuous treatment. Working Paper 32117, National Bureau of Economic Research.

Caves, K. and Lueling, S. (2021). Research and the real world: Analyzing the short- and long-term impact of knowledge transfer. *Research Evaluation*, 30(3):393–404.

Chowdhury, R. (2021). Grow the pie or have it? using machine learning for impact heterogeneity in the ultra-poor graduation model. *Center for Effective Global Action*.

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

DellaVigna, S., Kim, W., and Linos, E. (2022). Bottlenecks for evidence adoption. Working Paper 30144, National Bureau of Economic Research.

DHS Spatial Repository (2015). https://spatialdata.dhsprogram.com/data/#/. Online; accessed 26 April 2024.

Evans, D. and Acosta, A. (2021). What's the latest economic research on africa? a round-up of nearly one hundred studies from csae 2021 — cgdev.org. https://www.cgdev.org. Online; accessed 27-04-2024.

Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., and Vermeersch, C. M. J. (2016). *Impact Evaluation in Practice*. The World Bank.

Greenhalgh, T., Raftery, J., Hanney, S., and Glover, M. (2016). Research impact: a narrative review. *BMC Medicine*, 14(1).

Hoddinott, J., Maluccio, J. A., Behrman, J. R., Flores, R., and Martorell, R. (2008). Effect of a nutrition intervention during early childhood on economic productivity in guatemalan adults. *Lancet*, 371(9610):411–416.

Kremer, M., Gallant, S., Rostapshova, O., and Thomas, M. (2018). Is development innovation a good investment? evidence on scaling and social returns from usaid's innovation fund. Online; accessed 26 April 2024.

Lutje, V., Gerritsen, A., and Siegfried, N. (2011). Randomized controlled trials of malaria intervention trials in africa, 1948 to 2007: a descriptive analysis. *Malaria Journal*, 10(1).

Miguel, E. and Kremer, M. (2004). Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, 72(1):159–217.

Nakajima, N. (2020). Evidence-based policymaking in education.

Ndounga Diakou, L. A., Ntoumi, F., Ravaud, P., and Boutron, I. (2017). Published randomized trials performed in sub-saharan africa focus on high-burden diseases but are frequently funded and led by high-income countries. *Journal of Clinical Epidemiology*, 82:29–36.e6.

Nobel Prize (2019). The prize in economic sciences 2019. *The Nobel Prize*.

Ogden, T. (2020). *RCTs in Development Economics, Their Critics and Their Evolution*, page 126–151. Oxford University PressOxford.

Siegfried, N., Clarke, M., and Volmink, J. (2005). Randomised controlled trials in africa of hiv and aids: descriptive study and spatial distribution. *BMJ*, 331(7519):742.

State Department for Devolution (2024). www.devolution.go.ke/county-information. Online; accessed 27-04-2024.

Vivalt, E. and Coville, A. (2023). How do policymakers update their beliefs? *Journal of Development Economics*, 165:103121.

6 Appendix

6.1 Evidence of Parallel Trends

To test the parallel trends assumption, I collected data from 1998. Formally, the parallel trends assumption can be estimated using the following regression equation:

$$Y_{it} = \alpha + \beta_1 Treat_{1i} + \beta_2 Treat_{2i} + \beta_3 Post_t + \beta_4 Treat_{1i} \cdot Post_t + \beta_5 Treat_{2i} \cdot Post_t + \epsilon_i. \quad (2)$$

Here, $Post_t$ is a dummy variable that is equal to 1 if the observation is from 2003 and 0 if it is from 1998. The goal is to test whether β_4 and β_5 are statistically significant or not; ideally, they would be statistically insignificant, which would ensure that the control group is a valid counterfactual and thereby justifies the use of differences-in-differences model for this analysis.

Tables 7 and 8 provide the differences-in-differences estimates for Equation (2) for education and health variables respectively. None of the differences-in-differences estimates are statistically significant, except for the estimate for medium-intensity coverage in Panel A of Table 7. Hence, we can conclude that the parallel trends assumption holds and therefore the differences-in-differences design is justified.

Table 7: Parallel Trends Test for Education Variables

Panel A: EDEDUCMCPR

	(1)	(2)
Intercept	26.7***	27.7***
	(1)	(0)
Medium	3.6	-3.4**
	(4.32)	(1.85)
High	-6.8***	-1.1
	(1.56)	(7.9)
Post	-0.1	2.8***
	(4.03)	(0)
Medium x Post	-4.05	-6.17***
	(0.8)	(2.16)
High x Post	0.8	-2.7
	(4.2)	(10.04)
N	12	12
R^2	0.498	0.309
Adjusted R^2	0.080	-0.267
Residual Standard Error	26.95 (df = 6)	37.11 (df = 6)

Panel B: EDEDUCMCSC

	(1)	(2)
Intercept	23.1***	25.3***
	(2.2)	(0)
Medium	-0.35	-5.47***
	(4.01)	(0.62)
High	-2.25	-1
	(2.75)	(1.8)
Post	-6.45***	-7.7***
	(2.4)	(0)
Medium x Post	-0.35	1.83
	(5.13)	(1.18)
High x Post	-2.1	-1.45
	(3.08)	(4.25)
N	12	12
R^2	0.756	0.828
Adjusted R^2	0.553	0.684
Residual Standard Error	10 (df = 6)	7.07 (df = 6)

Panel C: EDEDUCWCPR

	(1)	(2)
Intercept	21.75***	21.8***
	(0.05)	(0)
Medium	7.1**	-0.47
	(3.15)	(3.041)
High	-2.65	5
	(2.5)	(5.2)
Post	1.15	0.18
	(1)	(0)
Medium x Post	2.6	4.3
	(3.45)	(4.61)
High x Post	0.25	-0.45
	(2.95)	(8.84)
N	12	12
R^2	0.877	0.208
Adjusted R^2	0.774	-0.451
Residual Standard Error	6.54 (df = 6)	42 (df = 6)

Panel D: EDEDUCWCSC

	(1)	(2)
Intercept	12.8***	12.2***
	(0.6)	(0)
Medium	4.35	0.67
	(3.4)	(0.86)
High	-1	4.15
	(0.72)	(4.15)
Post	-2.2**	-2.3***
	(0.92)	(0)
Medium x Post	-0.6	-0.93
	(5.26)	(1.67)
High x Post	-2.15**	-1.05
	(1.04)	(6.73)
N	12	12
R^2	0.662	0.375
Adjusted R^2	0.380	-0.147
Residual Standard Error	9.3 (df = 6)	17.2 (df = 6)

Note: The dependent variable in Panel A is the percentage of men who completed primary education. The dependent variable in Panel B is the percentage of men who completed secondary education. The dependent variable in Panel C is the percentage of women who completed primary education. The dependent variable in Panel D is the percentage of women who completed secondary education. *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively. The estimation is performed using OLS regression.

Table 8: Parallel Trends Test for Health Variables

Panel A: CHVACCCBAS

	(1)	(2)
Intercept	67.3***	67.7***
	(0.4)	(0.0)
Medium	2.4**	-7.7
	(1.17)	(9.01)
High	-19.5***	-6.7
	(3.423)	(9.8)
Post	-10.5**	-5.6***
	(5.315)	(0.0)
Medium x Post	5.85	-6.53
	(7.907)	(12.81)
High x Post	1.75	2.7
	(8.952)	(16.04)
N	12	12
R^2	0.866	0.264
Adjusted R^2	0.754	-0.350
Residual Standard Error	38.135 (df = 6)	208.922 (df = 6)

Panel B: CMECMRCCMR

	(1)	(2)
Intercept	67.3***	67.7***
	(0.4)	(0.0)
Medium	2.4**	-7.7
	(1.17)	(9.01)
High	-19.5***	-6.7
	(3.423)	(9.8)
Post	-10.5**	-5.6***
	(5.315)	(0.0)
Medium x Post	5.85	-6.53
	(7.907)	(12.81)
High x Post	1.75	2.7
	(8.952)	(16.04)
N	12	12
R^2	0.866	0.264
Adjusted R^2	0.754	-0.350
Residual Standard Error	38.135 (df = 6)	208.922 (df = 6)

Panel C: CNNUTSCHA2

	(1)	(2)
Intercept	32***	42***
	(10)	(0.0)
Medium	-11.5	8
	(14.5)	(27.3)
High	38.5	-15
	(28.32)	(17)
Post	-3	-9***
	(10.77)	(0.0)
Medium x Post	3	10.67
	(15.44)	(41.59)
High x Post	12.5	17.5
	(38.67)	(25.13)
N	12	12
R^2	0.681	0.107
Adjusted R^2	0.415	-0.637
Residual Standard Error	538.917 (df = 6)	1507.53 (df = 6)

Panel D: CNNUTSCWA2

	(1)	(2)
Intercept	82***	67.7***
	(14)	(0.0)
Medium	-26.5	-7.7
	(26.5)	(9.01)
High	78.5**	-6.7
	(40.97)	(9.8)
Post	14.5	-5.6***
	(24.01)	(0.0)
Medium x Post	-1	-6.53
	(36.16)	(12.81)
High x Post	0.0	2.7
	(54.95)	(16.04)
N	12	12
R^2	0.767	0.264
Adjusted R^2	0.572	-0.350
Residual Standard Error	1250.25 (df = 6)	4861.53 (df = 6)

Note: The dependent variable in Panel A is the percentage of children who received all 8 vaccinations. The dependent variable in Panel B is the child mortality rate. The dependent variable in Panel C is the percentage of children who are stunted. The dependent variable in Panel D is the percentage of children who are overweight. *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively. The estimation is performed using OLS regression.