THE ECONOMIC JOURNAL

The Economic Journal, **133** (*May*), 1348–1370 https://doi.org/10.1093/ej/uead001
Advance Access Publication Date: 12 January 2023

LONG-RUN EFFECTS OF AID: FORECASTS AND EVIDENCE FROM SIERRA LEONE*

Katherine Casey, Rachel Glennerster, Edward Miguel and Maarten Voors

We evaluate the long-run effects of a decentralised approach to economic development called community-driven development—a prominent strategy for delivering foreign aid—by revisiting a randomised community-driven development program in Sierra Leone 11 years after launch. We estimate large persistent gains in local public goods and market activity, and modest positive effects on institutions. There is suggestive evidence that community-driven development may have slightly improved the communities' response to the 2014 Ebola epidemic. We compare estimates to the forecasts of experts from Sierra Leone and abroad, working in policy and academia, and find that local policymakers are overly optimistic about the effectiveness of community-driven development.

Since the 1990s, community-driven development (CDD) has emerged as a dominant approach to distributing foreign aid to poor and vulnerable communities. At its core, CDD devolves control over the selection, implementation and management of local public goods to communities (White, 1999; Mansuri and Rao, 2013). This highly decentralised and participatory approach has two main goals: to bolster local public infrastructure and associated economic activity through the provision of block grants; and to democratise community decision-making via social facilitation focused on the inclusion of marginalised groups. Advocates see it as a particularly useful approach in post-conflict environments or where the state is weak (Wong and Guggenheim, 2018). As a leading donor, the World Bank alone spent $85 billion over the first two decades of CDD programming (Mansuri and Rao, 2013), and currently maintains $42.6 billion in active investments across 93 countries.[1]

Meta analysis of recent field experiments suggests that CDD effectively delivers local infrastructure, accompanied by little discernible impact on institutional outcomes, at least in the short run (Casey, 2018). There are almost no data on how CDD performs over the longer term. This is an important lacunae to fill in light of the often elusive nature of aid sustainability (Kremer and Miguel, 2007), and the open question of whether external reforms to strengthen

* Corresponding author: Katherine Casey, Graduate School of Business, Stanford University, 655 Knight Way, Stanford CA 94305, USA. Email: kecasey@stanford.edu

This paper was received on 19 July 2021 and accepted on 23 December 2022. The Editor was Ekaterina Zhuravskaya.

The data and codes for this paper are available on the Journal repository. They were checked for their ability to reproduce the results presented in the paper. The replication package for this paper is available at the following address: https://doi.org/10.5281/zenodo.7411565.

We thank the Decentralization Secretariat, the GoBifo Project, Local Councillors in Bombali and Bonthe districts and a panel of experts for their collaboration. We thank Naasey Kanko Arthur, Samuel Asher, Angélica Eguiguren, Erin Iyigun, Andrés F. Rodriguez, Mirella Schrijver, Eleanor Wiseman and the Innovations for Poverty Action team in Freetown for excellent research assistance and fieldwork. We thank Macartan Humphreys, Stefano DellaVigna, Eva Vivalt, numerous seminar participants and the 2018 BITSS Forecasting Conference for valuable comments. We gratefully acknowledge financial support from the UK Economic and Social Research Council, the Governance Initiative at J-PAL, NWO 451-14-001 and the Stanford Institute for Innovation in Developing Economies. Human subjects approval was obtained from the Sierra Leone Ethics and Scientific Review Committee, Stanford University (#38846), MIT COUHES (#1612798296) and Wageningen University. This study was pre-registered on the AEA registry: https://www.socialscienceregistry.org/trials/1784. All errors are our own.

[1] *Source*: https://www.worldbank.org/en/topic/communitydrivendevelopment#2 (last accessed 1 June 2022).

institutions can indeed succeed when afforded a sufficiently long time horizon. CDD offers an instructive application for these questions, given its policy prominence and commensurate resource allocation, as well as the fact that early programs are now 'ageing' into a stage where it is possible to assess longer-run effects (Bouguen *et al.*, 2019).

This study makes three contributions. First, it experimentally evaluates the impacts of a high-profile CDD aid program in Sierra Leone more than a decade after implementation began, using an array of measures to capture public goods provision, economic activity, social capital and local institutions. Second, it uses the 2014 Ebola public health crisis as a real-life test of the quality and adaptability of social capital and local institutions by assessing whether the CDD program enabled communities to better prepare for, and more effectively respond to, that crisis. And third, it compares program effects observed on the ground with the prior beliefs and forecasts of a large number of experts located in Sierra Leone and abroad, and working in both policy and academia.

The analysis centres on the 'GoBifo' CDD program,[2] which was implemented by the Government of Sierra Leone's Decentralisation Secretariat with support from the World Bank. A first intense phase of the program ran from 2005 to 2009, where treatment communities each received roughly $5,000 in block grants and six months of dedicated social facilitation. Participating communities established village development committees (VDCs), mandated to include representatives of marginalised groups, which were trained and encouraged to make the selection and implementation of local projects in a democratic manner. Program staff closely monitored community observance of these participation and inclusion rules, and both their administrative records and our survey data document widespread adherence. VDC members had the opportunity to learn by doing in managing a series of small-scale public projects funded by the grants, and liaised regularly with members of local elected government. A second less intense phase of the program commenced in 2010, which provided additional grants to a subset of treatment communities and continued some lighter touch engagement with project staff.

This is an informative context to study the long-run effects of CDD. The treatment was relatively intense, well implemented and impactful in the short run. In earlier work, we found substantial positive impacts on local public goods and economic activity, and stronger links between the community and local government, over the first four years of program activity (Casey *et al.*, 2012). Given the high rates at which aid-funded infrastructure has been found to fall into disrepair in similar contexts (Miguel and Gugerty, 2005), it is useful to assess whether public infrastructure provided under this CDD aid program persists, particularly as it is constructed at relatively low cost (Wong, 2012). It is further interesting to see whether a decade of strong national economic growth has bolstered the stock of local public goods and market activity, potentially helping control communities to catch up with their treated counterparts.[3]

Earlier work also found precisely estimated null results of CDD on a broad range of measures capturing institutional change, a finding that has since been challenged on both theoretical and econometric grounds, which provides a further motivation for a longer-term follow-up. Conceptually, some critics argue that the initial evaluation timeline may have been too short to capture impacts on slowly evolving institutions, especially if institutional change follows a non-linear trajectory (Woolcock, 2013). Statistically, Anderson and Magruder (2022) re-analysed

[2] 'GoBifo' means 'move forward' in the local Krio language.

[3] Annual gross domestic product (GDP) growth averaged 5.25% over the study period (World Bank, https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?end=2016&locations=SL&start=2005, last accessed 26 April 2022).

the earlier data using more flexible, and thus higher powered, econometric methods, and found support for positive short-run effects of CDD on multiple outcome measures dispersed across several hypotheses regarding institutional change. Partially in response to these perspectives, we returned (in 2016) to all 236 originally sampled communities, seven years after the short-run data collection (in 2009) and 11 years after the program launch (in 2005), in order to assess long-run changes in institutions, and evaluate the persistence of CDD investments in local public goods. We use the same (or very similar) indicators as in previous survey rounds in order to map changes over time.

Analysis of the long-run data uncovers strong persistence in the short-run gains observed for measures of infrastructural 'hardware', alongside smaller, but statistically significant increases, in indicators related to institutional 'software'. For hardware, we estimate the impacts of CDD on an index of 30 distinct outcomes relating to project implementation, local public goods and economic activity. We find a long-run treatment effect of 0.204 SD units (SE 0.040), which is two-thirds the magnitude of what we estimated in the short run. Given the difficult post-conflict operating environment and high levels of asset depreciation, we view these results as encouraging. For institutions, the estimated treatment effect on an index of 63 measures capturing collective action, inclusion, trust, groups and networks (among others) is 0.059 SD units (SE 0.024). While this point estimate is precise and larger than what we observe in the short-run data, it is small in magnitude.[4]

Another dimension of social capital and institutions is the extent to which it allows a community to cope with unanticipated shocks. The 2014 outbreak of the Ebola virus disease in West Africa is the largest ever recorded, and the crisis resulted in over 4,000 deaths in Sierra Leone alone (of roughly 11,000 in total in the broader region). Some of the actions the government asked communities to take to prepare for and respond to cases—such as create community by-laws, report suspected cases and disseminate prevention information—could be facilitated by local institutional capacity of the kind GoBifo aimed to build, which our experimental design enables us to evaluate. We find suggestive evidence for small-sized positive treatment effects on a subset of indicators relating to community actions (as opposed to knowledge or health practices). These results on the impact of CDD on the communities' ability to respond in a crisis are consistent with our finding of small positive impacts of CDD on social capital and institutions more generally. Our results are also consistent with evidence from a contemporaneous study that found that previous community mobilisation efforts led to more effective Ebola responses in this same empirical context (Christensen *et al*., 2021).

We elicited the prior beliefs of experts about the prospects for long-run change, and compare their predictions to our empirical estimates. We collected these data in 2016–7, which, to our knowledge, is among the first such elicitations for a field experiment. This enables us to assess the accuracy and variability of well-informed forecasters in this context. Specifically, we collected priors regarding the long-run effects of CDD aid on both institutional and infrastructural outcomes from 126 experts familiar with CDD, a group that includes practitioners in Sierra Leone and multilateral institutions, like the World Bank, as well as research faculty in economics and political science, and their graduate students.

[4] In a companion paper, we leverage a separate experiment that we overlaid across this study sample to compare the effectiveness of CDD to a more technocratic alternative that identifies residents with high human capital, and encourages communities to put them in charge of development projects (see Casey *et al.*, 2021).

Here we find wide dispersion in the prior beliefs of domain experts about the scope for long-run change, particularly with regards to institutional performance, which makes the 2016 data collection an interesting empirical exercise. One striking pattern that emerges across outcomes is the consistently more optimistic view towards this type of foreign aid among Sierra Leonean policymakers, in contrast to the overall pessimism among academic researchers. While it is too early to tell exactly when and how such predictions will be most useful, this exercise adds a few data points to broader efforts to systematically document prior beliefs, and compare them to outcomes obtained in lab and field settings (see DellaVigna and Pope, 2018; Vivalt and Coville, 2020; Vivalt *et al.*, 2021; among others).[5]

1. Material and Methods

1.1. *Intervention and Research Timeline*

The 236 communities in Sierra Leone tracked over an 11-year period are located in two districts, Bombali and Bonthe (see Figure 1). They were selected to balance regional diversity, political affiliation and ethnic composition, while simultaneously targeting poor rural areas that had previously received little aid. Half of these communities were randomly assigned to participate in the GoBifo CDD program and the remaining half to the control group that received no assistance. Baseline data were collected in 2005 before program activity commenced.

The program hired facilitators to help treatment communities assemble a VDC, which was required to include both women and young men (both considered marginalised groups). Facilitators then trained VDC members how to select, plan, implement and monitor local development projects in an inclusive and democratic way. The first and most intense phase of GoBifo (2005 to 2009) disbursed roughly $5,000 per treatment community, or approximately $100 per household, for use in constructing small-scale public goods (like latrines, community centres and cement floors for drying agriculture produce) or enterprise support (like training and start-up capital for carpentry and garment dying).

During weekly visits, GoBifo staff conducted training, facilitated meetings and tracked participation in program activities. Accumulated over the course of the first few years of the program, these visits and training translated into six months of dedicated in-person support per community. The objective was to permanently lower the fixed cost of collective action—which could make future inclusive community decisions and development activities easier—and thereby place communities on a stronger development trajectory that would outlast the direct financing stage.

To capture short- to medium-run impacts, the research team collected data in 2009 on 12 hypotheses about how CDD could alter community outcomes. Three of these hypotheses concern the 'hardware' of development, like public goods provision and economic activity, and the remaining nine capture measures of institutional 'software', like social capital, inclusion and participation (see Online Appendix Table A2 for a detailed list of study hypotheses).[6] These hypotheses were developed in partnership with the CDD practitioner team in 2005. Casey

[5] A platform has been established to collect these forecasts systematically; see DellaVigna *et al.* (2019) and https://socialscienceprediction.org/.

[6] To give some examples of outcomes under each family, development hardware includes measures of project implementation (like the establishment of a village development committee), enumerator inspection of a standard suite of local public infrastructure items and measures of market activity (like the number of goods available for purchase). Software includes measures of trust, membership in social groups, community meetings, conflict and knowledge of governance processes. See Online Appendix Table A5 for a complete list of outcomes.

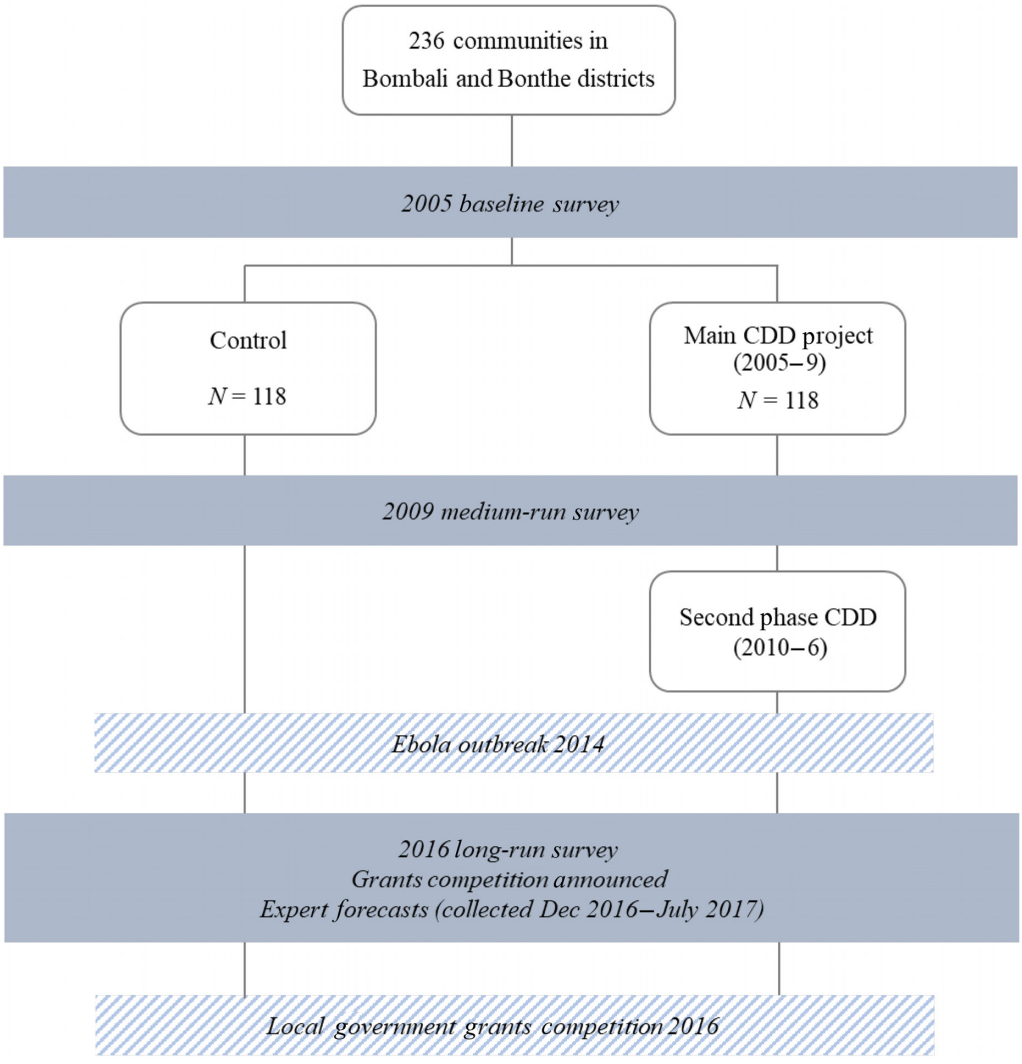

Fig. 1. *Research Design and Timeline.*
Notes: CDD treatment assignments are displayed in rounded boxes, research activity and data collection in shaded rectangles, and external events and activities in hashed rectangles.

et al. (2012) analysed the 2009 data and found strong positive impacts for the hardware family of outcomes, and a series of precisely estimated null results for the software family. This empirical pattern broadly resonates with other short-run experimental studies of CDD programs in Afghanistan (Beath *et al.*, 2013), the Democratic Republic of Congo (Humphreys *et al.*, 2019) and Liberia (Fearon *et al.*, 2015).

A less intensive phase of GoBifo began in 2010. The program disbursed additional grants to 60 of the 118 treatment communities, amounting to $1,300 per community to support youth empowerment activities ('youth' is defined by the government as individuals under 35 years of

age).[7] Once again, no activities were implemented in the control communities. Facilitation staff in the two district headquarters (as well as management staff in the capital) were employed full time throughout this second period, and remained on the government payroll at least through the long-run data collection, in 2016. They continued some project facilitation activities in treatment villages, although we lack reliable data on the frequency of these interactions, and our impression is that, beyond the grants noted above, the level of operational support for treatment villages was minimal after 2012.

In 2016, field enumeration teams returned to the original sample of 236 villages in order to collect long-run data, covering both the original 12 research hypotheses as well as a new hypothesis about community responses to the 2014 Ebola epidemic. Analysis in this paper thus evaluates the persistence of the initial financial and organisational investments made under the first intense phase of GoBifo, plus any additional effects of the subsequent treatment 'dose' delivered in the second phase. Total project costs for the first phase (2005–9) are approximately $2 million, and for the second, less active phase (2010–8) nearly $3 million, given the continuation of project staffing, transport and overhead for several years. Thus, from a broader policy perspective, we evaluate a $5 million investment in CDD that was at least nominally operational for more than a decade.

1.2. *Long-Run Data Collection*

The 2016 long-run data collection aimed to replicate as closely as possible the infrastructure and institutional measures collected in 2009, as well as extend consideration to new measures capturing community responses to the Ebola crisis. To do so, field teams conducted focus group discussions with local leaders, and physically inspected a suite of community amenities and observable indicators of market activity. Note that, while the 2009 data collection included both household- and community-level surveys, budgetary constraints limited the 2016 collection to community-level outcomes only. Where possible, we include community-level analogues of unmeasured household-level indicators; however, the set of indicators collected in 2016 remains a subset of that collected in 2009. We pre-registered all outcomes and analysis in the AEA registry (see https://www.socialscienceregistry.org/trials/1784 and the pre-analysis plan in Online Appendix D).

We supplement survey indicators with directly observed communal behaviours. This exercise aims to loosely replicate the structured community activities (SCAs) that we developed in 2009 and discussed in Casey *et al.* (2012). In 2016, we measured whether and how communities responded to a project challenge competition that the elected district councils were running at the time. This community competition awarded a total of 20 grants worth $2,500 each to support local public infrastructure projects, selected based on the quality of proposals submitted by communities. To publicise this opportunity, supervisors of data collection teams held a public meeting in all study communities. Supervisors explained that to enter the competition communities needed to develop a project idea and complete a standardised, but somewhat technical three-page proposal. They then asked community members to nominate five people who had the requisite skills to lead the community through the proposal process. The enumeration teams then stood back outside the meeting and allowed communities to deliberate as they saw fit. Enumerators discretely observed

[7] This subset of 60 of the treatment communities was not randomly selected. Sponsored activities included the provision of school or sports materials (e.g., uniforms, classroom materials), training for small-scale entrepreneurs (like tailoring), construction (e.g., drying floors, toilets, storage, school buildings) and farming implements.

the ensuing proceedings and recorded information on how the deliberation unfolded, the presence and engagement of youth and women, and the influence of local leaders on the process. These measures of observed behaviour expand and deepen our analysis of local institutional inclusion and performance.

1.3. *Expert Prior Elicitation*

To assess whether the results of the 2016 data collection were in line with expert priors, we asked knowledgeable policymakers and academics to make a series of predictions before we analysed any of the data. Experts were asked to make forecasts in three areas: the long-run effects of CDD on (*i*) infrastructure and (*ii*) institutions, and (*iii*) the response of communities to the district government grants competition.

The first two categories of forecasts (infrastructure and institutions) were at the heart of the Casey *et al.* (2012) study, and we therefore structured the data collection and expert forecasts around the same 12 hypotheses used in earlier work. For each hypothesis, the survey instrument restates the hypothesis (e.g., 'Hypothesis 1: GoBifo Project Implementation'), provides an example of indicators used to measure the hypothesis (e.g., 'Examples of indicators include the presence of a village development committee and formal bank account for village project expenses'), and asks for a prediction about the long-run results using a slider bar that ranges from −0.50 to +0.50 SD units (see the instrument in Online Appendix A). As not all experts are familiar with this metric, the survey describes what SD units are and provides rules of thumb for what constitutes small versus large effects. We randomly varied whether or not the survey prompted the expert with the medium run results about CDD (e.g., 'our study found medium-run effects for this hypothesis equal to +0.20 sdu's, which is statistically different from zero with a very high degree of confidence').

We then asked experts about the grants competition. This section of the survey provided background information on the competition and the procedures the field supervisors followed in publicising it to communities, including the process for generating nominations for local residents who could lead the proposal process. Finally, we asked for predictions about what percentage of communities would enter the competition.[8]

A broad variety of experts participated, including those in Sierra Leone. Through systematic outreach we collected priors from 126 experts, including policymakers in Sierra Leone with knowledge of the GoBifo project; policy experts working for multilateral aid agencies such as the World Bank, primarily based in Organization for Economic Co-operation and Development (OECD) countries; faculty in both economics and political science who have been involved in evaluating CDD projects or related areas of development (including all the authors of this article); and economics students in Sierra Leone (undergraduates) and OECD countries (doctoral students). Survey response rates were relatively high for all groups (e.g., 84% for faculty and 99% for graduate students) except for OECD policymakers (39% completion). We discuss potential implications of this differential response rate in the results section below. The variety of experts surveyed allows us to test whether policymakers and academics have similar levels of optimism and/or bias regarding intervention impacts, and assess regional differences in perspective between international experts versus those located in the host country.

[8] This section of the expert survey also references a new experiment to promote delegation of project management to high skill local residents that we overlaid on top of this long-run CDD sampling frame, which is analysed in Casey *et al.* (2021).

1.4. *Empirical Strategy*

To assess the long-run impacts of CDD, we estimate the model

$$Y_c = \beta_0 + \beta_1 CDD_c + \boldsymbol{W}_{\boldsymbol{c}}'\Psi + \boldsymbol{X}_{\boldsymbol{c}}'\boldsymbol{\Gamma} + \varepsilon_c, \quad (1)$$

where outcome Y (e.g., presence of a public good, institutional outcome or Ebola response measure) is measured for each community c; *CDD* is an indicator for participation in the GoBifo program; W_c is a vector of stratification fixed effects for geographic wards; the X_c are balancing variables used in the original 2005 randomisation (community size and distance to the nearest road) and ε_c is an idiosyncratic error term.[9] We further test for heterogeneous treatment effects along the same eight community-level dimensions we used (and measured) in the short-run analysis (namely, total households, war exposure, average schooling, distance to the road, historical domestic slavery, district, ethnic fractionalisation and chiefly authority; see Online Appendix Table A3).[10]

Throughout the analysis, we adjust for the fact that we conduct multiple tests on the same dataset by implementing false discovery rate (FDR) corrections (see Benjamini *et al.*, 2006 and Anderson, 2008). These adjustments run across the two outcome families, or across all 12 individual hypotheses, as relevant. We also report the 'naïve' or 'per-comparison' p-value for those interested in a particular hypothesis on its own.

We test directly for decay in the estimates from the short to long run using the model

$$Y_c^L - Y_c^S = \gamma_0 + \gamma_1 CDD_c + \boldsymbol{W}_{\boldsymbol{c}}'\Psi + \boldsymbol{X}_{\boldsymbol{c}}'\boldsymbol{\Gamma} + \varepsilon_c,$$

where the dependent variable is the difference in the mean effect indices measured in the long-run survey, Y_c^L, and the short-run survey, Y_c^S. All other variables remain as defined above for (1). The coefficient of interest is γ_1, where $\gamma_1 < 0$ suggests that the treatment effect dissipated over time on average for that outcome. Note that the set of outcomes varies between the 2009 and 2016 data collection rounds, so each index incorporates the relevant outcomes for that particular survey round (see Online Appendix Table A1 for estimates limited to the exact panel outcomes).

2. Results

2.1. *Long-Run CDD Effects on Infrastructural Hardware Outcomes*

We find evidence for positive, highly significant impacts of the CDD program on measures of development hardware over the long run. For the overall 'family' of infrastructure outcomes, Table 1, panel A reports a long-run treatment effect of 0.204 SD units, which is sizeable in magnitude and statistically significant at the 99% confidence level. Estimates do not change substantively when we limit the set of outcomes to those that form an exact panel (which includes 29 of the original 39 outcomes from 2009): the 2016 treatment effect estimate is 0.208 SD units (SE 0.041) in Online Appendix Table A1.

This positive effect reflects gains across the three component hypotheses: project implementation (e.g., does the community have a VDC?); the stock and quality of local public infrastructure (e.g., does the community have a functional water well?) and economic activity (e.g., how many

[9] The original randomisation was successful in generating treatment and control groups that are well balanced on observable characteristics; see Online Appendix Table A8.

[10] Consistent with Casey *et al.* (2012), we find little evidence for heterogeneous effects, save for smaller impacts in one of the two study districts, namely the Bombali district.

Table 1. *Long-Run CDD Treatment Effects.*

	Treatment effect 2016 (1)	Naïve *p*-value (2)	FDR *q*-value (3)	Treatment effect 2009 (4)	Change over time (1) − (4)
Panel A: infrastructure 'hardware' family					
All outcomes (30 unique outcomes)	0.204***	<0.001	0.001	0.298***	−0.094***
	(0.040)			(0.031)	(0.036)
Project implementation	0.253***	<0.001	0.001	0.703***	−0.450***
	(0.067)			(0.055)	(0.080)
Local public goods	0.228***	<0.001	0.001	0.204***	0.024
	(0.046)			(0.039)	(0.041)
Economic welfare	0.240***	<0.001	0.001	0.376***	−0.136**
	(0.056)			(0.047)	(0.062)
Panel B: institutions 'software' family					
All outcomes (61 unique outcomes)	0.066***	0.009	0.005	0.028	0.038
	(0.025)			(0.020)	(0.028)
Collective action	0.098**	0.050	0.099	0.012	0.086
	(0.050)			(0.037)	(0.061)
Inclusion	0.036	0.338	0.291	0.002	0.034
	(0.038)			(0.032)	(0.045)
Local authority	−0.050	0.380	0.296	0.056	−0.106
	(0.056)			(0.037)	(0.069)
Trust	0.107*	0.065	0.109	0.042	0.064
	(0.057)			(0.046)	(0.081)
Groups and networks	0.149**	0.038	0.094	0.028	0.121
	(0.071)			(0.037)	(0.074)
Access to information	−0.036	0.591	0.476	0.038	−0.075
	(0.067)			(0.037)	(0.072)
Participation in governance	0.079	0.191	0.194	0.090***	−0.011
	(0.060)			(0.045)	(0.065)
Crime and conflict	−0.002	0.971	0.480	0.01	−0.012
	(0.063)			(0.043)	(0.074)
Political and social attitudes	0.154	0.216	0.194	0.041	0.113
	(0.124)			(0.043)	(0.126)
Observations	236				

Notes: Significance levels are represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ and based on naive per-comparison values. Specifications include strata for geographic ward and two balancing variables (distance to the road and community size) from the original randomisation. All estimates are for hypothesis-level equally weighted mean effect indices, expressed in SD units (see Kling *et al.,* 2007). Column (3) includes *q*-values from FDR corrections across the 12 hypotheses and across the two family indices, respectively (see Benjamini *et al*., 2006 and Anderson, 2008). The dependent variable in column (5) is the difference in the 2009 and 2006 indices, where the set of component measures varies across survey rounds (see Online Appendix Table A1 for an exact panel specification). The 2009 data are sourced from Casey *et al*. (2012).

goods are for sale in the community?). For each hypothesis, the CDD treatment effect estimate is positive and large in magnitude, ranging from 0.228 to 0.253 SD units. The estimates are highly statistically significant (in column (2)), even after accounting for the fact that we are testing multiple hypotheses on the same dataset (in column (3)).

Decay over time is moderate: the family-level long-run effect of 0.204 SD units is two-thirds the size of the effect estimated in the short run, which was 0.298 SD units (in column (4)). This suggests a considerable degree of persistence, even years after most direct financial support ceased. The estimated decay of roughly one-third of the original effect is statistically distinct from zero (column (5)).

Project implementation exhibits both the largest estimated short-run effect (of 0.703 SD units in column (4)) and the strongest decay over time (of −0.450 SD units in column (5), or roughly

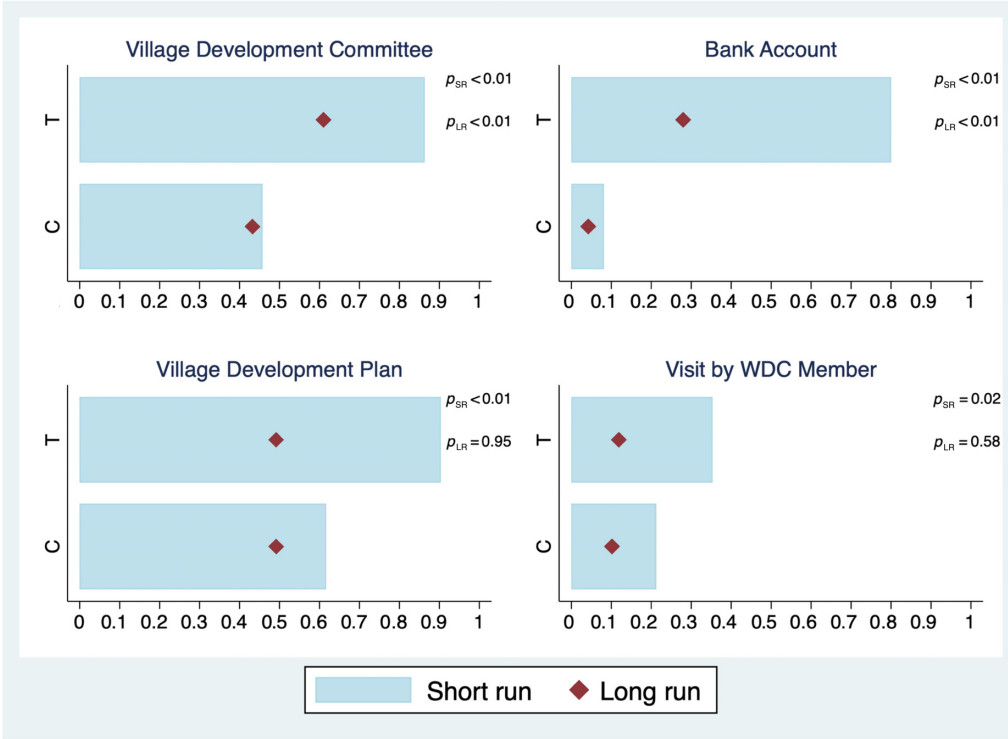

Fig. 2. *Decay in Project Implementation Effects.*
Notes: This figure compares short- and long-run treatment effect estimates for an illustrative sample of outcomes under the first hypothesis concerning project implementation. The solid horizontal bars denote the mean proportion of communities in the treatment (top) versus control (bottom) group that register the presence of a given outcome in the short-run 2009 data. The diamonds indicate the corresponding proportions observed in the long-run 2016 data. Reported *p*-values are associated with treatment effect estimates that include the full suite of pre-specified controls in the short-run (denoted p_{SR}) and long-run (denoted p_{LR}) datasets, respectively. All four outcomes are measured as binary indicators.

half) in this family. To provide a clearer sense of the magnitude of these effects, consider a few of the underlying outcomes presented in Figure 2 (see Online Appendix Table A5 for treatment effect estimates for all individual outcome measures). The solid bars denote the proportion of communities in treatment and control groups, respectively, with the particular outcome observed in the short-run 2009 data, while the diamonds reflect the corresponding proportions in the long-run 2016 data. For the presence of a VDC, we see that in the short-run data, treatment communities were more likely to have a VDC by about 40 percentage points (on a base rate of 46% in control communities). In the long run, the prevalence of a VDC in control communities remained roughly constant (at 43% in 2016), while the treatment effect fell to a 17 percentage point difference.

The pattern for the other two core CDD operational measures in Figure 2 (establishing a community bank account and drafting a village development plan) looks similar, with large gains in the immediate aftermath of the project that have strongly dissipated over time. This is consistent with the 'project bubble' conjecture (Wong, 2012), whereby the organisational architecture

established by CDD was leveraged effectively during project implementation, but then not repurposed or used for much subsequent local development activity outside the CDD sphere. For the fourth outcome in this set (visits by a member of the Ward Development Committee (WDC)), the relatively more modest short-run gains have fully dissipated by 2016, further suggesting that the attention from and connection to public officials facilitated by the CDD project approval process did not translate into enduring relationships between participating communities and this most local tier of elected government.

In contrast to what we see for project implementation, there is no statistically detectable change in treatment effects from the short to long run for the second hypothesis about impacts of the program on the stock and quality of local public goods (Table 1, panel A, third row). The estimated treatment effect on the index of outcomes in 2009 was 0.204 SD units (SE 0.039), compared to 0.228 (0.046) in 2016. At the level of individual outcomes, this effect captures durable improvements in the availability of functional agricultural drying floors, traditional birth attendant huts and court 'barries' (or public buildings for dispute resolution), among others.

Figure 3 shows the persistence in the increased number of distinct local public goods (out of a standardised set of 12 infrastructure items) that were deemed present and functional by field enumerator inspection.[11] The solid bars show the number of public goods observed in the short run, with the diamonds denoting the corresponding count observed in the long run. This stock, as well as the positive wedge created by the CDD program, has held fairly steady over time for communities in the treatment and control groups. On average, control communities exhibited 2.7 (2.5) out of the 12 standard items, compared to 3.7 (3.5) for treatment communities, in the short (long) run. (See Online Appendix Figure A1 for each infrastructure item broken out on its own.) This suggests very little depreciation in these public assets over time. It is further worth noting that the mean number of public goods remains low for both groups, indicating high levels of deprivation in these communities throughout the study period.

Measures of economic welfare (hypothesis 3) suggest that one-third of the initial gains dissipated over time, from an estimated treatment effect of 0.38 SD units (SE 0.05) in the short run to 0.24 SD units (0.06) in the long run (in the fourth row of Table 1). This reflects persistent increases in local market activity, including enumerator observation of petty traders active and the number of common items available for purchase in the community on the day of the 2016 field visit.

Figure 4 graphs this observed pattern of decay using an illustrative selection of economic welfare outcomes. As above, the solid horizontal bars represent the average number (out of five) economic indicators measured in 2009 for treatment and control separately, and the diamonds reflect the 2016 counts. The underlying set of variables includes the presence of petty traders, an above-median number of goods for sale, a field enumerator assessment that the village is better off than others in the area, an above-median number of new businesses started in the past two years and an above-median number of small shops in the community.[12] Unlike above, economic welfare appears to be increasing over time for both groups (the count in controls moves from 2.3 to 3.0, and for treatment, from 2.8 to 3.3) and control communities appear to be closing the initial CDD-induced gap. (See Online Appendix Figure A2 for each indicator on its own.)

[11] These infrastructure measures span 12 of the 17 total outcomes grouped under this second hypothesis. See Online Appendix Table A5 for treatment effect estimates for all individual outcomes.

[12] This graph omits three outcomes under this hypothesis: (*i*) bank account, as it is already displayed in Figure 2; and (*ii*) two others (time since the most recent visit by an external trader and participation in skills training) that are not measured in the same fashion across the two data collection rounds. See Online Appendix Table A5 for results for all individual outcomes.

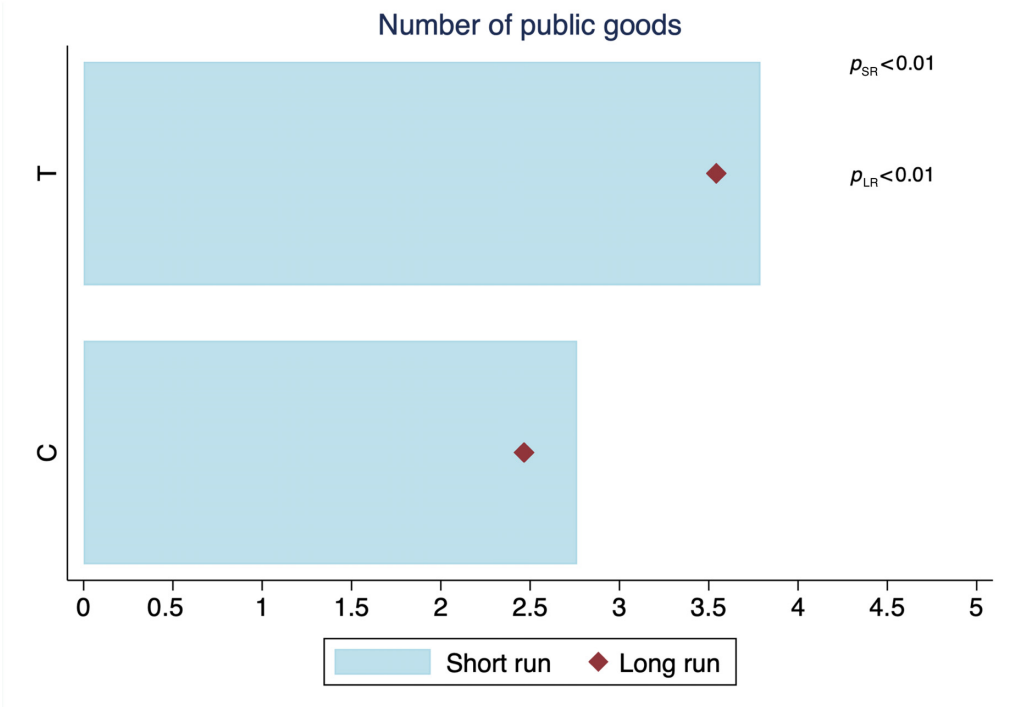

Fig. 3. *Persistence of Public Infrastructure Gains.*
Notes: This figure displays the average number of functional public goods, out of a standardised set of 12 local infrastructure items, observed in the community by field enumerator inspection. The solid horizontal bars denote the average count for treatment (top) versus control (bottom) communities observed in the short-run 2009 data. The diamonds indicate the corresponding counts observed in the long-run 2016 data. Reported *p*-values are associated with treatment effect estimates that include the full suite of pre-specified controls in the short-run (denoted p_{SR}) and long-run (denoted p_{LR}) datasets, respectively. (For treatment effect estimates for each individual public good, see Online Appendix Figure A1.) The standardised list of local public goods includes a primary school, public health unit, water well, agricultural drying floor, grain store, community centre, 'palava' hut (or conflict resolution site), court barrie (or court structure), market, latrine, traditional birth attendant hut and sports field.

Thus, a decade of strong national economic growth appears to have a broad buoying effect that is narrowing the estimated treatment effect over time.

In our view, these results showing persistent gains in the 'hardware' family of development outcomes are impressive, and particularly so, given the challenges of working in a post-conflict environment. These family-level gains further appear to be widely distributed across treatment communities (as opposed to concentrated in a few high performing areas). To see this, Figure 5(a) plots the kernel density distribution of the total number of outcomes under this infrastructure family that are observed in treatment (thick solid line) and control (thin solid line) communities over time. Each individual outcome is expressed as a binary indicator with '1' indicating more favourable status (e.g., the presence of a particular item, or an above-median value of a continuous variable), and we include all 29 outcomes that form an exact panel over the two survey waves.

In the short run, control communities exhibit a roughly normal distribution of outcomes under the infrastructure family with a median of nine outcomes present (left-hand graph). For the

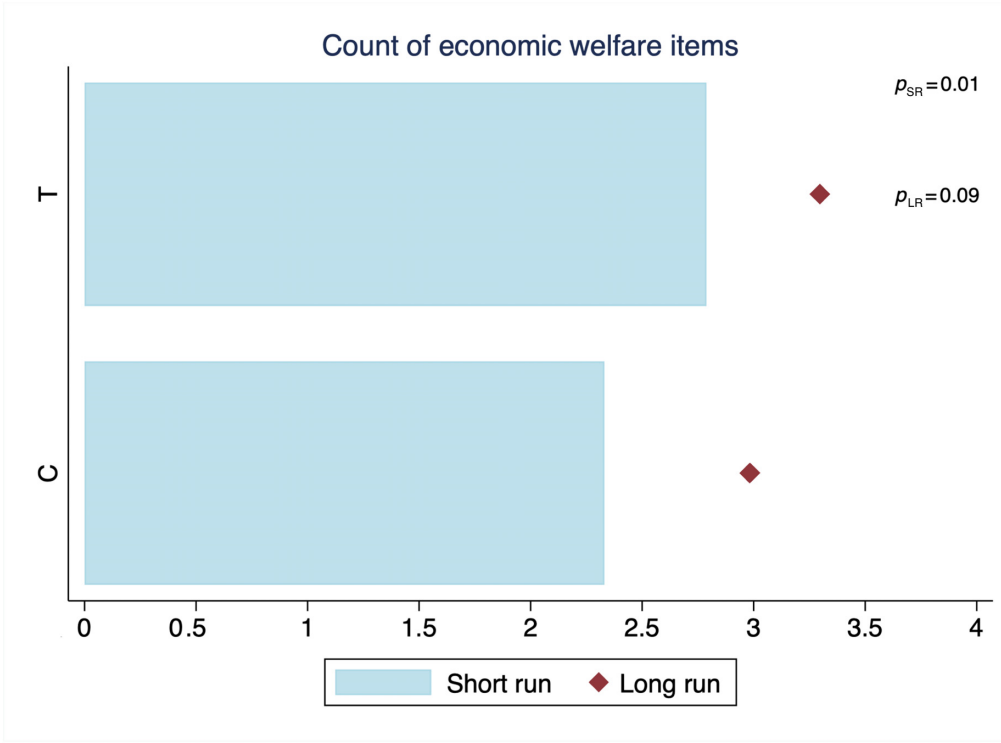

Fig. 4. *Economic Catch Up by Control Communities.*
Notes: This figure highlights the relatively strong gains in measures of economic welfare over time observed for study communities. The solid bars display the average count of five binary indicators of economic welfare for treatment (top) versus control (bottom) communities observed in the short-run 2009 data. The diamonds indicate the corresponding counts observed in the long-run 2016 data. Reported *p*-values are associated with treatment effect estimates that include the full suite of pre-specified controls in the short-run (denoted p_{SR}) and long-run (denoted p_{LR}) datasets, respectively. (For treatment effect estimates for each individual economic welfare item, see Online Appendix Figure A2.) The set of indicators includes the presence of petty traders, number of goods for sale, field enumerator assessment that the village is better off than others in the areas, number of new businesses started in the past two years and the number of small shops in the community. Continuous outcomes are transformed into binary indicators that equal one if the observation is above the median value.

treatment group, the distribution is shifted to the right across the full range with a median of 13 outcomes present. A Kolmogorov–Smirnov test rejects the equality of these distributions at the 99% confidence level. The distribution of treatment group outcomes in the long run remains shifted to the right of the control distribution, but the differences are smaller (right-hand graph). The respective median values are now 10 for control and 12 for treatment. A Kolmogorov–Smirnov test rejects the equality of these long-run distributions at the 99% confidence level.

While the experimental design does not allow us to directly compare infrastructure provision under CDD versus other delivery mechanisms, there are some useful benchmarks in the literature. Miguel and Gugerty (2005), for example, found that nearly half of borehole water wells built by a European bilateral aid donor in Kenya in the 1980s were no longer functional within a decade of construction. Our estimated loss for the hardware family overall is only one-third for

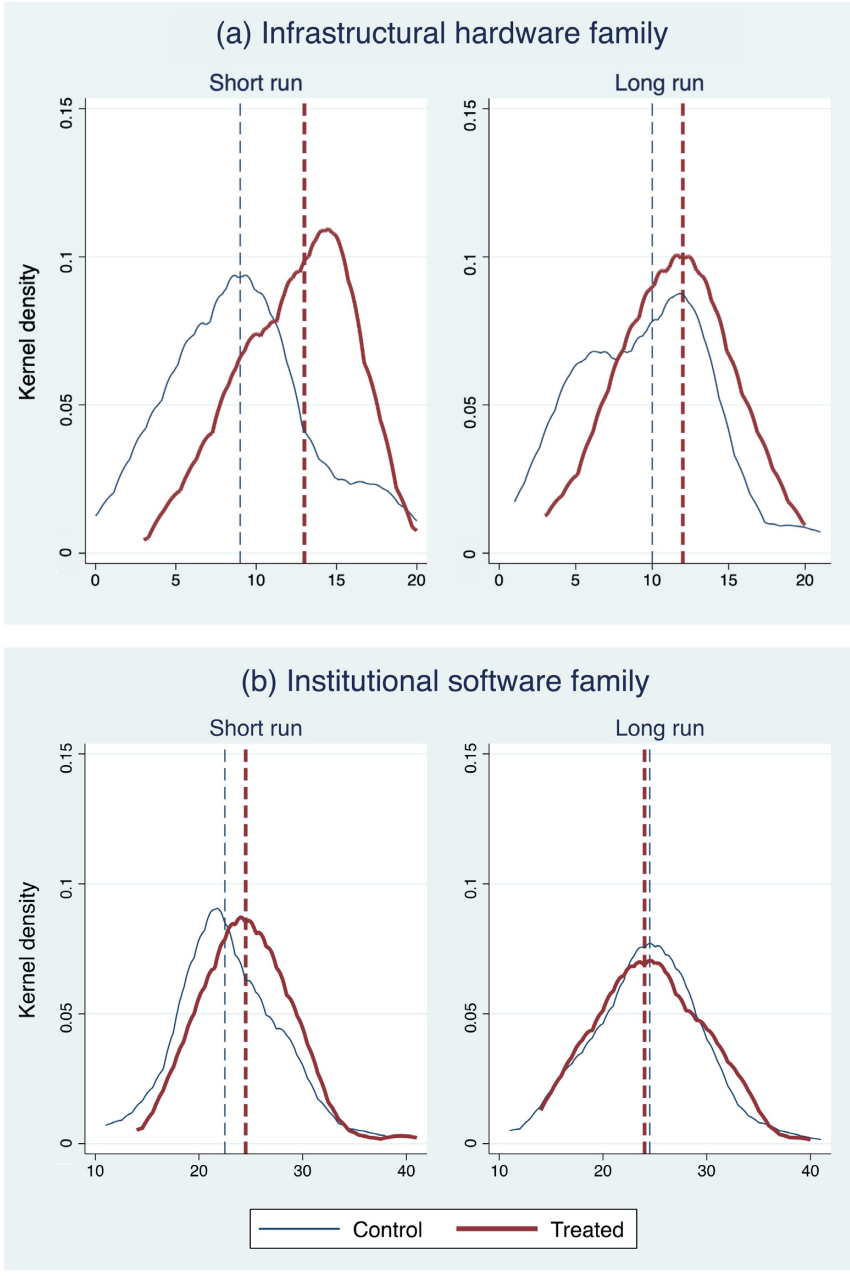

Fig. 5. *Distribution of Treatment Effects.*

Notes: This figure plots the kernel density distribution of outcomes for treatment and control groups in the short- and long-run data. Panel (a) focuses on the 'hardware' family of 29 outcomes, while panel (b) presents the 'software' family of 57 outcomes. Both sets are limited to outcomes that form an exact panel over time. The *y* axis in each plot demarcates the number of indicator variables registering a favourable outcome, with continuous outcomes transformed into binary indicators that equal one if the observation is above the median value. Dashed vertical lines demarcate the median number of outcomes observed in treatment and control samples separately.

CDD investments over a comparable time frame. The comparatively strong CDD performance is particularly encouraging, given that CDD projects tend to be implemented at lower cost than other government service delivery mechanisms (Wong, 2012), raising the question of whether they were done to a lower standard. While we cannot parse mechanisms underlying the CDD effect, these relatively favourable results are at least consistent with CDD advocates' claims about the value of local participation in aligning investments with demand, and thereby bolstering utilisation and maintenance over time (Dongier *et al.*, 2002). Relatively simple local construction practices may also make routine maintenance easier.

The Sierra Leone results provide evidence for stronger positive effects when compared to the one other longer-run CDD experiment that we are aware of, namely Mvukiyehe and Van der Windt (2020) in the Democratic Republic of Congo (DRC). While they found some positive effects for the persistence of physical infrastructure, they estimate null results for long-run impacts on service delivery, economic welfare, social inclusion and local institutions. Our study is distinct from theirs in that it operates over a longer time horizon (returning 11 versus 8 years after the project launch) and follows up on stronger short-run results (see Humphreys *et al.*, 2019 on null results for the DRC program), which provides a more relevant setting for investigating the persistence of effects.

2.2. *Long-Run CDD Effects on Institutional Software Outcomes*

Analysis of the 2016 data yields small positive estimates for the long-run effects of CDD on local institutions. Combining all 61 individual outcomes grouped under this family into an equally weighted index yields a positive, precisely estimated, but small in magnitude treatment effect of 0.066 SD units (SE 0.025) in Table 1, panel B. Of the nine distinct hypotheses about how CDD might alter institutions, three are positive (collective action, trust, and groups and networks), and at least marginally significant on a per-comparison basis (column (2)). Precision decreases somewhat when adjusting for multiple inference (column (3)). One way to interpret this pattern of results is that if we conceive of all outcomes measuring a latent variable associated with institutional quality, CDD had a small positive impact, but the effect is not large enough to clearly parse effects along the nine underlying channels.

To illustrate what is improving in the three areas where we estimate non-zero hypothesis-level treatment effects, we present results for some of the underlying component variables. For collective action, the two largest positive estimates are located on indicators that share commonality with the hardware family: this includes the presence of a VDC, which is directly cross-listed under both families; and the presence of a communal farm, which could result from CDD-funded agricultural projects. Among the components of the trust hypothesis, we estimate significant increases for trusting non-governmental organizations (NGOs) and people from outside your community, which suggests that the experience interacting with GoBifo staff left an enduring positive impact on community perceptions of outsiders. Finally, for groups and networks, the estimated treatment effect is positive in sign for eight of the nine distinct groups enumerated (e.g., credit and savings, parent teacher association (PTA), seed multiplication, religious groups, etc.), which is consistent with a broad strengthening of associational ties among community members. See Online Appendix Table A5 for estimates for all individual outcomes.

To address potential concern about bias in self-reports elicited in the community survey, we focus on directly observed behaviours, using the 11 SCA measures relating to the project

challenge application. The overall treatment effect in Online Appendix Table A4 is 0.009 with SE 0.057. On a per-comparison basis, only one individual indicator registers a statistically significant effect, which is a large positive effect on the time that the community took to generate its list of five nominees. In a companion paper, we analyse a broader array of outcomes related to the grants competition, and find weak evidence for CDD effects on intermediate measures—like the village chief's willingness to delegate proposal authority to one of the community nominees—but null results for the ultimate outcomes of interest, which is the quality of the proposal and ultimately the probability of winning one of the actual grants (see Casey *et al*., 2021).

How does the small positive long-run effect on institutions compare to what was measured in the short run? While the 2016 point estimate is more than twice as large in magnitude as the null result for 2009 (0.066 versus 0.028 SD units), the estimated decay over time is not statistically distinct from zero (in column (5)). Yet, recall that these two estimates operate over different subsets of indicators, as the long-run data collection does not include household surveys.[13] If we limit consideration to outcomes that were collected in identical fashion across the two survey rounds, the overall CDD treatment effect remains the same for 2016 (at 0.065, SE 0.026, in Online Appendix Table A1). The 2009 effect, however, is somewhat larger and becomes statistically significant (at 0.086 SD units, SE 0.030). This increase in the 2009 effect could reflect differences in reporting between households and community leaders (although it is unclear to us ex ante which group is more or less susceptible to social desirability bias), or could be due to sampling variation created by focusing on a subset of outcomes. The fact that the relative magnitude of the short- versus long-run effect varies across these two specifications, while the estimated coefficient on decay is not significant in either, implies that we cannot say anything definitive about the dynamics of institutional change over time. It is clear, however, that the estimated magnitudes are modest and broadly similar across specifications, lying in the 0.028 to 0.086 SD unit range.

To explore the distribution of institutional outcomes across communities, Figure 5(b) plots the kernel density of these measures in the same intuitive count fashion as discussed above. We limit consideration to the 57 (of 61 total) outcomes that form an exact panel over time, again transforming all outcomes into binary indicators (using an above-median cut for continuous variables). In the left-hand graph, the short-run distribution of total favourable outcomes observed for treatment communities lies to the right of that for controls at lower levels, and has a somewhat higher median (24.5 versus 22.5 outcomes). A Kolmogorov–Smirnov test rejects equality of these short-run distributions at the 95% confidence level (associated p-value of 0.033). The long-run data displayed in the graph on the right show very closely overlapping graphs, with almost identical medians (24 versus 24.5 outcomes), and a Kolmogorov–Smirnov fails to reject equality of the long-run distributions (associated p-value of 0.95). This is in contrast to the exact panel regression estimate in Online Appendix Table A1, where we find a small yet significant treatment effect. The difference between the regression results and the Kolmogorov–Smirnov test is due to the transformation of continuous outcomes into binary indicators used in the figure. The binary transformation inevitably cannot pick up impacts on the tails of the continuous variable and has less power. Nevertheless, the sensitivity of the significance of the institutional result to such

[13] Compared to the infrastructure family, which is based primarily on enumerator assessments of physical goods in both rounds, the lack of household data matters more here. Specifically, the 2009 round paired all community-level indicators (e.g., a count of how many people are observed at a particular community meeting) with reports from representative households (e.g., did any member of this household attend this particular meeting?), so excluding the household reports from the 2016 round cuts the number of institutional measures by roughly one-half.

transformations again underlines the conclusion that the impacts on institutional software are relatively small in magnitude.

Looking across families, it is interesting to note that communities that perform well in one domain also tend to perform well in the other. In the full sample, the unconditional correlation between the hardware and software family-wise indicator counts (as displayed in Figure 5) is 0.613 in the endline data. Honing in on the treatment group, changes from the 2005 baseline to 2016 endline are also positively associated across families, with a correlation coefficient of 0.347 (see Online Appendix Figure A3).[14] While speculative, this finding suggests that CDD-induced institutional changes allowed communities to achieve greater infrastructural gains over the long run.

Given the extended time horizon of the study, one natural concern is that spillovers from treatment to control communities, perhaps through learning or migration, are leading to an underestimate of the long-run treatment effects. While we do not have direct data on this (recall that we did not survey households in the 2016 round), we believe spillovers are unlikely to play a substantial role for a few reasons. To start, Figure 5(a) shows that the count of hardware indicators in control communities has held fairly steady over the years, as opposed to increasing as one would expect if gains in treatment communities were being shared with controls. This is intuitive, given the fixed location and substantial up-front costs associated with infrastructure creation. For institutional software, the count of indicators in panel (b) has indeed increased over time for controls; however, the gains are small. Moreover, for specific institutional practices that would be relatively easy to learn and import into control communities, Figure 2 shows no gains over time in practices like establishing a VDC or formulating a development plan.

2.3. *CDD Impacts during a Public Health Crisis*

While measuring the same indicators over time allows us to track the persistence of impacts, we also investigated how prepared communities were to deal with the unprecedented shock of the Ebola public health crisis of 2014. CDD could strengthen the Ebola response through two mechanisms: the higher quantity and quality of infrastructure could have been utilised to directly respond to the epidemic, or stronger institutions (including improved collective action and greater trust of NGOs) could have enabled more rapid and effective behavioural responses. Note that CDD resources were not sufficient to pay for the construction of a new clinic, and we see no difference between CDD and non-CDD communities in the presence of health clinics, which suggests that the institutional channel may be more relevant (see Online Appendix Table A5).[15] Analysis in this section covers a variety of related outcomes, such as the creation of an Ebola task force and knowledge about the epidemic (on symptoms, transmission and control).

The estimated treatment effect for CDD on the index of the 13 combined Ebola knowledge items and response actions is small in magnitude, and not statistically distinguishable from zero (0.042 SD units, with an SE of 0.036 in Table 2). For the Bombali district, which was harder hit by Ebola, the effect is also null (−0.001, 0.053, $N = 156$ communities; see Online Appendix Table A6), while for the Bonthe district, it is positive and statistically significant (0.109, 0.053, $N = 80$).

[14] There are fewer outcomes that form an exact panel with the 2005 baseline data: 12 (24) for family A (B).

[15] While CDD did positively impact the presence of traditional birth attendant huts, which are in the healthcare realm, these maternal huts are less likely to be relevant for the Ebola response.

Table 2. *CDD Treatment Effects on Ebola Knowledge Items and Response Actions.*

Outcome	Mean, controls	Treatment effect	Standard error	*p*-value	FDR *q*-value
Mean effects index (all 13 indicators)	0.000	0.042	0.036	0.249	–
Knowledge items					
Mean effects index (all nine knowledge items)	0.000	0.015	0.050	0.766	0.621
Correctly answers 'No' to 'Can Ebola spread through air?'	0.856	−0.005	0.040	0.896	0.999
Correctly answers '21' to 'How many days can it take for the first to symptoms arise?'	0.669	0.014	0.051	0.791	0.999
Total (of 11 possible) correct answers to questions about how one can get Ebola	5.220	0.006	0.187	0.974	0.999
Knows correct Ebola hotline number	1.000	0.000	0.000	–	–
Total (of 10 possible) correct answers regarding how to protect yourself against Ebola	4.975	−0.051	0.201	0.801	0.999
Correctly answers 'No' to 'Drinking salt water can help cure Ebola?'	0.958	0.030	0.019	0.114	0.999
Correctly answers 'No' to 'Drinking chlorine can help cure Ebola?'	1.000	−0.009	0.009	0.321	0.999
Correctly answers 'No' to 'Can someone spread Ebola before they show symptoms?'	0.695	0.030	0.052	0.565	0.999
Total correct answers (of 14 possible) regarding symptoms of Ebola	7.263	−0.230	0.232	0.324	0.999
Response actions					
Mean effects index (all four response actions)	0.000	0.090*	0.053	0.091	0.223
Community had an Ebola task force during the Ebola crisis	0.661	0.077	0.052	0.145	0.999
Community created by-laws in relation to Ebola	0.907	0.042**	0.019	0.030	0.563
Communities are more likely to go to formal health facilities (nurse, clinic)	0.924	0.014	0.030	0.632	0.999
Communities are more likely to go to formal health facilities for Ebola (nurse, clinic)	0.915	0.000	0.034	0.995	0.999
Observations	236				

Notes: Significance levels are based on naive *p*-values and represented by * $p < 0.10$, ** $p < 0.05$. Specifications include strata for geographic ward and two balancing variables (distance to the road and community size) from the randomisation. This table includes 13 of 15 pre-specified primary outcomes in our pre-analysis plan (PAP), excluding two outcomes that are observed for fewer than 20 communities in the data. The *q*-values are from FDR corrections adjusted across the two category-level indices, or across all 13 individual knowledge and response outcomes, respectively.

For individual outcomes, while we find no change in measures regarding community members' knowledge of Ebola, we do see some evidence that communities had taken more action. At the bottom of Table 2, communities were 4% more likely to have established Ebola related by-laws (significant at 95% confidence, a small increase compared to the high level of compliance in control villages of 91%) and 8% more likely to have established an Ebola task force (not significant, up from 66% in controls). In a combined index of all response actions, we find a small, positive and marginally significant effect of 0.090 SD units (SE 0.053) in the full sample. None of these estimates remain significant after adjusting for multiple inference (column (5)).

Taken together, this provides suggestive evidence that the CDD program may have generated some benefits for villages during the Ebola crisis, although the effect magnitudes are small and concentrated in the category of response action indicators. This pattern of results for Ebola is broadly consistent with what we saw above for institutional impacts, where analysis detected small gains clustered under three of the nine hypotheses in the family. The comparison to

institutional impacts is intuitive, given that the support and actions requested are similar in nature: for example, communities received no financial support to respond to Ebola and were asked primarily to engage in voluntary collection action.

These results are less pronounced than, although in the same direction as, findings from Christensen *et al*. (2021), who analysed the impact of a community mobilisation program in Sierra Leone to increase accountability, and trust in local health clinics through facilitated community monitoring and engagement.[16] The program was implemented before the Ebola outbreak. Just prior to the crisis, they found that the program interventions built confidence in health workers and improved the perceived quality of care. During the crisis, this led to more reporting of Ebola cases and lower mortality from the disease (conditional on the case being reported). Our weaker results may reflect the fact that Gobifo was not specifically targeted at community mobilisation in public health, but taken together, the two papers raise the possibility that community mobilisation may be an effective strategy to generate collective action under crisis conditions.

2.4. *Expert Forecasts*

There is much debate about the role of experts in international development and whether their opinions are informative or helpful (Easterly, 2014), and relatedly whether knowledge of what has worked elsewhere is useful in understanding what might work in a given context. More prosaically, it is useful from a scholarly perspective to understand if the results of a study are a surprise, or are in line with previous expectations. Our results suggest that, collectively, experts do on average have useful knowledge, but their opinions are highly variable.

Figure 6 displays forecasts for three distinct areas: CDD effects on infrastructure (panel (a)), CDD effects on institutions (panel (b)) and community entry into the grants competition (panel (c)). For each type of expert (e.g., policymaker or academic faculty), the open circles represent individual expert predictions, and the filled circle denotes the mean prediction for the group with a whisker plot displaying the accompanying 95% confidence interval. We compare these forecasts to the realised effect size estimated in the 2016 data, which is denoted with a solid horizontal line, with dashed lines demarcating the 95% confidence interval.

Starting with long-run impacts on CDD-funded infrastructure investments, pooled together, the experts predicted a long-run treatment effect of 0.218 SD units (SE 0.126), which is statistically indistinguishable from the estimated effect (of 0.204). There is wide dispersion in forecasts—ranging from zero to 0.5 SD units—which is evident both within and across the different types of expert. Policymakers in Sierra Leone were relatively more optimistic about persistent infrastructure gains and faculty more pessimistic. The predictions of economics students generally track those of policymakers in their respective regions.

For institutions, experts in Sierra Leone were particularly optimistic about the scope for long-run impacts (panel (b)). Policymakers and students alike in Sierra Leone predicted average effects in the range of 0.25 SD units, which turned out to be a substantial overestimate compared to the realised effect size (of 0.066). Policymakers and students in the OECD on average were roughly on target. While we cannot reject that economics and political science faculty were correct on average, they were more pessimistic: a substantial number of them (11 out of 23) predicted precisely zero long-run effects, which falls outside the 95% confidence interval of the observed

[16] The program consisted of an additional treatment arm that provided non-financial recognition for clinic staff. Results across both arms were similar, though generally stronger for the community monitoring intervention.

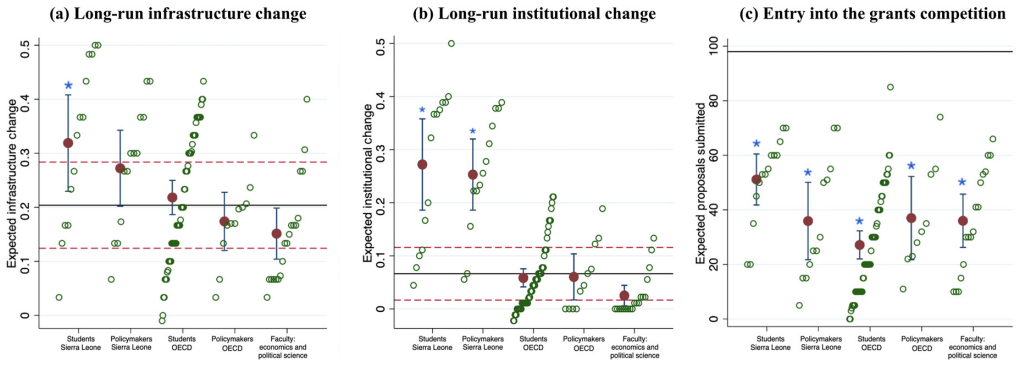

Fig. 6. *Expert Predictions of Long-Run CDD Effects and Grants Competition.*
Notes: This figure presents expert predictions collected during December 2016 and July 2017 before any data analysis. Panels (a) and (b) present expectations for CDD treatment effects measured in SD units. The realised effect size is denoted with a solid horizontal line and the accompanying 95% confidence interval is demarcated by dashed horizontal lines. Panel (c) presents expectations about the percentage of CDD control communities that would enter the grants competition. The realised point estimates are the 0.204 SD unit CDD treatment effect for infrastructure in panel (a); the 0.066 SD unit CDD treatment effect on institutions for panel (b) and 98.3% of communities entered the grants competition for panel (c). For panels (a) and (b), expert predictions were closer to the realised value for the version of the survey that provided the short- to medium-run results for institutional change (p-value < 0.01), but not statistically distinct for infrastructure (p-value $= 0.27$). Stars above the 95% confidence interval denote forecasts that are significantly different from the realised effect.

point estimate.[17] If we pool all expert predictions together, the long-run forecast for institutional change significantly exceeds what was estimated in the short run (0.095 predicted by experts, compared to 0.028 units in Casey *et al.*, 2012). This difference remains statistically distinct from zero, even when limited to the subgroup of experts who were randomly chosen to be primed with additional information on the short-run results (results not shown).

The substantial ex ante disagreement among seemingly well-informed experts about CDD's long-run institutional impacts, makes the 2016 data collection an interesting empirical exercise, and particularly so in light of the accumulation of shorter-run null results for institutional outcomes from several studies (see Wong, 2012; King and Samii, 2014; White *et al.*, 2017 and Casey, 2018 for cross-country reviews). Moreover, the divergence between policymakers in Sierra Leone and academics lends some credence to concerns about optimism bias among policymakers and gripes (from policymakers) about hard-to-please academics, although note the substantial variation in priors among both types of expert. This potential disconnect does not appear to be as severe for policymakers based in the OECD countries, suggesting that the feedback loop between academic results and policy perceptions may be working relatively well for policymakers who are more proximate to rich country scholars, perhaps due to more frequent interactions at conferences and policy fora.

By contrast, all expert opinion diverged substantially from observed outcomes regarding entry into the infrastructure grants competition. As a group, the experts predicted a baseline take up rate of 36% for control communities, which reflects the sentiment of one expert who cautioned

[17] The co-authors of this paper, whose forecasts are excluded from Figure 2, predicted more pessimistic long-run outcomes with an average of 0.147 (SD 0.144) for hardware outcomes and 0.008 (SD 0.017) for software outcomes.

that 'it is very likely that $2,500 is just too small an amount to get enough communities to bother with applying'. In practice, we found a take up rate of 98%, which surprised all experts and far exceeded any prediction in the sample (in panel (c)). Online Appendix Table A7 shows that experts on average expected CDD treatment communities to take up the grants opportunity at slightly higher rates than controls (by 7 percentage points), a difference that we do not observe in practice.

There are a couple of potentially important differences across the distinct pools of experts, which may confound some of the apparent divergence in their predictions, necessitating some caveats about this exercise. First, the survey response rate was markedly lower for OECD policymakers as compared to the other four groups. If only the most interested or knowledgeable OECD policymakers completed the survey, their greater accuracy (at least for panels (a) and (b)) may be more attributable to positive self-selection, or greater effort in formulating their responses, than a generalisable difference between these policymakers and other types of expert. Second, formulating accurate statistical predictions is challenging in general, and may be particularly so for experts without formal statistical training. Policymakers in Sierra Leone are less likely to have such training than their OECD counterparts (in both policy and academia), and the students surveyed in Sierra Leone were undergraduates, whereas those in the OECD were doctoral level, which may have contributed to the domestic experts' overestimation when using SD units (again in panels (a) and (b)). We do not have enough data to systematically parse these channels, and believe that this offers a promising avenue for future expert predictions (similar to the work by Vivalt *et al.*, 2021).

3. Discussion and Conclusion

Community-driven development commands a substantial share of foreign aid allocations and is particularly common in post-conflict situations. Its short-run effects have been fairly extensively studied by randomised controlled trials in several different countries. This study broadens the evidence base by (*i*) extending the time horizon to capture longer-run effects than any existing study in the literature (to our knowledge), (*ii*) evaluating impacts during a subsequent public health crisis and (*iii*) comparing expert forecasts to observed impacts.

First, following up with communities more than a decade after baseline data collection, we document strong persistence of CDD aid impacts on measures of development hardware, commensurate with two-thirds of the short-run gains (measured seven years prior). The pattern of decay differs across the three component hypotheses. For project implementation outcomes, like the presence of a VDC, the large short-run gains have strongly dissipated over time, suggesting that such organisational architecture was not re-purposed for much post-CDD development activity. By contrast, there is no statistically significant decay in the stock and quality of local public goods created by CDD, consistent with the program's emphasis on participation as a tool to align investments with local demand, and thereby foster local ownership and maintenance. This is an encouraging result, and particularly so in light of the difficult operating environment and the low cost of the infrastructure grants. And for economic welfare, while a strong decade of national growth has boosted outcomes in both treated and control communities, those that benefited from CDD still remain ahead in the long-run data.

Second, we find modest positive long-run effects on local institutions, which runs contrary to our own prior beliefs, although it seems unlikely that these small effects (+0.066 SD units

on average) are of major practical consequence. Similarly, we find suggestive evidence that the program may have helped communities respond slightly more effectively to the 2014 Ebola epidemic. While it is too early to understand the effect that GoBifo may have had on community preparedness and outcomes during the ongoing (at the time of writing) COVID-19 pandemic, nor do we have the data to do so, this finding from the Ebola crisis opens the possibility, at least speculatively, that earlier CDD programming may translate into positive gains during crisis.

Finally, comparing the empirical estimates to expert forecasts, we find wide dispersion in prior beliefs, a high degree of accuracy for some types of experts on particular outcomes, accompanied by systematic over- or underestimation for others. Taken together, the forecasts offer a few data points on the question of when and how expert predictions may be useful in research: we see (*i*) wide dispersion of views regarding the durability of infrastructure, (*ii*) disagreement across expert type for institutional change and (*iii*) systematic underestimation for community entry into the grants competition. One striking pattern is the consistent optimism regarding this type of foreign aid among Sierra Leonean policymakers, in contrast to the overall pessimism among researchers. This could be problematic if their sanguine view of institutional change drives the continued popularity of CDD programming. If, by contrast, policymakers are primarily motivated by the positive infrastructural effects, this would be less of a concern.

While expert prior opinions may be useful for predicting some effects, but not others, it remains unclear (to us) how to distinguish these cases ex ante. As more studies collect prior beliefs about the efficacy of policy interventions, a practice that is gaining some traction, the research community will be able to build a more thorough understanding of what types of impacts experts can reliably predict, and which types of experts—those with country knowledge, for instance, or practitioner experience or academic training—are most accurate.

Stanford University, USA & NBER, USA
University of Chicago, USA & NBER, USA
University of California, Berkeley, USA & NBER, USA
Wageningen University & Research, The Netherlands

Additional Supporting Information may be found in the online version of this article:

Online Appendix
Replication Package

References

Anderson, M. (2008). 'Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedaian, Perry preschool, and Early Training projects', *Journal of the American Statistical Association*, vol. 103(484), pp. 1481–95.

Anderson, M.L. and Magruder, J. (2022). 'Highly powered analysis plans', Working Paper 23544, National Bureau of Economic Research.

Beath, A., Christia, F. and Enikolopov, R. (2013). 'Do elected councils improve governance? Experimental evidence on local institutions in Afghanistan', Research Paper 2013-24, MIT Political Science Department.

Benjamini, Y., Krieger, A.M. and Yekutieli, D. (2006). 'Adaptive linear step-up procedures that control the false discovery rate', *Biometrika*, vol. 93, pp. 491–507.

Bouguen, A., Huang, Y., Kremer, M. and Miguel, E. (2019). 'Using RCTs to estimate long-run impacts in development economics', *Annual Review of Economics*, vol. 11, pp. 523–61.

Casey, K. (2018). 'Radical decentralization: Does community driven development work?' *Annual Review of Economics*, vol. 10, pp. 139–65.

Casey, K., Glennerster, R. and Miguel, E. (2012). 'Reshaping institutions: Evidence on aid impacts using a pre-analysis plan', *Quarterly Journal of Economics*, vol. 127(4), pp. 1755–812.

Casey, K., Glennerster, R., Miguel, E. and Voors, M. (2021). 'Skill versus voice in local development', *Review of Economics and Statistics*, doi: 10.1162/rest_a_01082.
Christensen, D., Dube, O., Haushofer, J., Siddiqi, B. and Voors, M. (2021). 'Building resilient Health Systems: Experimental evidence from Sierra Leone and the 2014 Ebola outbreak', *Quarterly Journal of Economics*, vol. 136(2), pp. 1145–98.
DellaVigna, S. and Pope, D. (2018). 'What motivates effort? Evidence and expert forecasts', *The Review of Economic Studies*, vol. 85(2), pp. 1029–69.
DellaVigna, S., Pope, D. and Vivalt, E. (2019). 'Predict science to improve science', *Science*, vol. 366(6464), pp. 428–9.
Dongier, P., Domelen, J.V., Ostrom, E., Rizvi, A., Wakeman, W., Bebbington, A., Alkire, S., Esmail, T. and Polski, M. (2002). 'Community-driven development', in *Core Techniques and Cross-Cutting Issues*, vol. 1, pp. 301–31, Washington, DC: World Bank.
Easterly, W. (2014). *The Tyranny of Experts: Economists, Dictators, and the Forgotten Rights of the Poor*, New York: Basic Books.
Fearon, J., Humphreys, M. and Weinstein, J. (2015). 'How does development assistance affect collective action capacity? Results from a field experiment in post-conflict Liberia', *American Political Science Review*, vol. 109(3), pp. 450–69.
Humphreys, M., Sánchez de la Sierra, R. and Van der Windt, P. (2019). 'Exporting democratic practices: Evidence from a village governance intervention in Eastern Congo', *Journal of Development Economics*, vol. 140, pp. 279–301.
King, E. and Samii, C. (2014). 'Fast-track institution building in conflict-armed countries? Insights from recent field experiments', *World Development*, vol. 64, pp. 740–54.
Kling, J.R., Liebman, J.B. and Katz, L.F. (2007). 'Experimental analysis of neighborhood effects', *Econometrica*, vol. 75(1), pp. 83–119.
Kremer, M. and Miguel, E. (2007). 'The illusion of sustainability', *Quarterly Journal of Economics*, vol. 122(30), pp. 1007–65.
Mansuri, G. and Rao, V. (2013). 'Localizing development: Does participation work?', Policy Research Report, World Bank.
Miguel, E. and Gugerty, M.K. (2005). 'Ethnic diversity, social sanctions, and public goods in Kenya', *Journal of Public Economics*, vol. 89, pp. 2325–68.
Mvukiyehe, E. and Van der Windt, P. (2020). 'Assessing the longer term impact of community-driven development programs evidence from a field experiment in the Democratic Republic of Congo', Policy Research Working Paper 9140, World Bank.
Vivalt, E. and Coville, A. (2020). 'How do policy-makers update their beliefs?', Working paper, University of Toronto.
Vivalt, E., Coville, A. and Sampada, K.C. (2021). 'Weighing the evidence: Which studies count?', Working paper, University of Toronto.
White, H. (1999). 'Politicizing development? The role of participation in the activities of aid agencies', in (K.L. Gupta, ed.), *Foreign Aid: New Perspectives*, vol. 68, pp. 109–25, Boston, MA: Springer.
White, H., Menon, R. and Waddington, H. (2017). 'Community-driven development: Does it build social cohesion or infrastructure? A mixed-method evidence synthesis report', 3ie Working Paper.
Woolcock, M. (2013). 'Using case studies to explore the external validity of 'complex' development interventions', *Evaluation*, vol. 19(3), pp. 229–48.
Wong, S. (2012). 'What have been the impacts of World Bank community-driven development programs? CDD impact evaluation review and operational research implications', Working paper, World Bank.
Wong, S. and Guggenheim, S. (2018). 'Community-driven development: Myths and realities', Policy Research Working Paper 8435, World Bank.