

When Does “What Works” Work?

And What Does that Mean for UK Aid R&D Spend?

Charles Kenny, Euan Ritchie, and Lee Robinson

The UK’s Secretary of State for International Development¹ oversees an aid-financed R&D² budget that is larger than that of the next 15 biggest donors combined.³ At the moment, a considerable proportion of that UK R&D spend goes towards solving global technological challenges related to neglected tropical diseases including malaria, and a considerable proportion again towards local evaluation of aid-financed development interventions. Much of the rest is somewhat opaquely distributed to British universities for research supposedly related to development. As well as reform of this last category, the range of more legitimate activities benefiting from ODA “research and development” calls for innovation in approaches to deliver outcomes.

This paper will argue there is a (fuzzy) spectrum of development procedures, for some of which global innovation, evaluation, or “best practice” can be informative, for some of which local evaluation or experimentation can be useful, and for some of which perhaps only practical experience and local wisdom can help. That there is a spectrum of intervention types and research opportunities, and that local evidence is often required, has implications for the kind of research that UK aid can usefully support as part of its R&D program and where that research should happen. In turn, that suggests a reform agenda for the way UK ODA for R&D is currently spent.

TYPES OF PROCEDURE AND TYPES OF KNOWLEDGE

Think of a spectrum regarding the nature of a “procedure” from the “technological” invention (creating a vaccine that, if administered correctly, protects against measles) to the culturally or environmentally “embedded” (rollout of a campaign to persuade people to abandon genital cutting). Add a second dimension running from procedures at the individual level (a vaccination) to the national level (monetary policy).

1 Thanks to Pam Jakiela for comments on an early draft of part of this material.

2 Robinson, Ritchie, and Kenny, “UK Research Aid: Tied, Opaque, and Off-Topic?”

3 This according to data submitted to CRS

A technological procedure is one that “works” similarly everywhere. An embedded procedure is one for which outcomes are determined by a range of environmental conditions: institutions, culture, infrastructure, factor endowments, and so on. At the limit, the complete set of conditions necessary for an embedded procedure to deliver its intended outcome is too big to be practically knowable.

Depending where procedures rest on the technological-embedded and individual-global scale, the kind of knowledge that is likely to be useful (or even attainable) regarding efficacy is likely to vary. The creation and testing of a vaccine against a particular strain of a particular disease is (often) something that only needs to happen once to give reasonably strong external validity to the global population (i.e., if it works here and now, it will work there and then). The external validity of evidence regarding a procedure that involves embedded institutions or interacts with culture is likely to be more limited (the Walmart logistics system won’t work for a kiosk in Mali). The ability to rigorously evaluate procedures also varies—it is easier at the individual level (pills for diarrhoea) than at the national level (industrial policy for growth).

This model (presented in Figure 1) is too simple. Not least, there are many intermediate positions and intermediate approaches to knowledge generation; somewhere between “best practice” and “practical local wisdom” lies local wisdom informed by the experience of others, for example. Also, some procedures that theoretically could be selectively applied at the individual level will (should) never be in practice, on human rights grounds, for example (access to a jury trial?). And in fact, scale effects and spillovers may change the impact of individual approaches tried at a larger scale.

Individual procedures as a matter of practicality are a combination of the technological and the cultural, the embedded and the indivisible. Take vaccines—in order to have an impact the technology needs to be rolled out in a place where the disease for which it provides prophylaxis is present, using a vaccination campaign that reaches beneficiaries at the right time with a nondegraded product, and where beneficiaries are willing to get vaccinated.

It is important to note “what works” in terms of procedures is not the only type of question worth asking with aid-funded R&D. Asking “what is” may often inform a more appropriate use of resources. Precisely because of the embedded nature of many procedures, knowing about the current conditions of a community or economy will be crucial to determining what is worth trying. Local knowledge and perspectives—especially from those often excluded from decision-making—should affect what interventions are proposed.

Figure 1. What Knowledge is Useful?

		Scale of Procedure (spectrum)	
		Individual	(Inter)national
Nature of Procedure (spectrum)	Technological	Information from anywhere	“Best practice”
	Embedded	Local information	Practical (local) wisdom

TOP LEFT: “INVENTED” TECHNOLOGIES

The global public good created by the inventors of technologies including vaccines, antibiotics, oral rehydration therapy, dwarf crop varieties, the internal combustion engine, the dynamo and turbine, plastic sheeting, the mobile phone, TCP/IP, the LED bulb, batteries, and the radial tire are immense. These technologies have proven themselves useful all over the world—and perform similarly enough all over the world that poorer people and poor countries spend at least as much, if not more, of their incomes purchasing these technologies as do rich people and countries. The great majority of these technologies were invented in the rich world before spreading to developing countries, although that is not true of many of the applications to which these technologies have been put—mobile finance, for example.

Realising the real-world impact of such technologies requires more than a single laboratory invention process. Even where a vaccine has already been created, the cost of additional research work in adapting it to the strains of microorganisms that exist in different climates can sometimes be considerable. Research and development of vaccines themselves had to be combined with (the perhaps surprisingly portable model of) vaccine delivery to have an impact. Vaccination rates for diphtheria, pertussis, and tetanus in Afghanistan were 24 percent⁴ in 2000 compared to around 65 percent today, for example. Partly as a result, child mortality in Afghanistan has approximately halved⁵ over the past 20 years. But Afghanistan has also seen vaccination workers attacked and vaccine rollouts challenged thanks to concerns over vaccine safety. Having an effective vaccine is only the first step.

In other cases, applications including mobile finance appear to be less easily portable—more “embedded”—than the underlying technologies. And it is worth emphasizing that “what is primarily techno-

4 World Bank, “World Development Indicators: Immunization.”

5 World Bank, “World Development Indicators: Under 5 Mortality.”

logical” may not be obvious (at least to outsiders). So, for example, why cookstove design turns out to be immensely culturally fraught whereas similar cell phone models appear to work worldwide was perhaps obvious to anthropologists and certainly obvious to those who cooked but was not necessarily obvious to inventors—or at least it appears that this came as some surprise to a number of cookstove innovators.⁶

More broadly, Comin and Mestieri look at a number of productive technologies including railways, telephones, ships, and spindles to suggest that the intensity of use of those technologies is strongly related to income per capita.⁷ That may suggest the return to productive technology use (especially if embedded in considerable capital like spindles in a textile factory) can be lower where infrastructure, human capital, or institutions are weak, even if they “work” the same everywhere. For example, “productive technology” may simply not be the binding constraint. Tomatoes in Kano in Nigeria rot in fields because roads aren’t good enough to transport them to processing plants in Lagos. Using more “productive technology” would add to the waste, as it doesn’t target the right problem. This in turn would suggest that approaching the solution to income divergence requires not just technology transfers, but adapting technologies to be better fits for developing countries and to ensure they have requisite institutions and infrastructure.

TOP RIGHT? “UNIVERSAL” MACRO POLICIES

There are numerous procedures and policy approaches that do appear to have worked across country contexts. That monetary policy worldwide has managed to almost completely eradicate the scourge of very high inflation⁸ and spiralling black market premiums on currency is a real success⁹ of the last 30 years.

Mobile telecommunications competition also seems to work pretty much everywhere. We don’t have a good natural experiment when it comes to the impact of telecoms reform, and Ethiopia¹⁰ retained a government-owned monopoly in mobile communications while seeing access increase pretty dramatically, but a fair number¹¹ of studies¹² have suggested faster sector growth after reform.¹³ Of course, that the technical solution “works” does not solve questions of political plausibility. Producing a blueprint for private competition in telecoms may be a technical exercise, but even technical reform creates winners and losers, and the latter are wont to mobilise in defence of their interests.

Technical areas where there is a broadly accepted blueprint for approaches are the natural home of technical assistance and “best practice” guidelines. The role for R&D is somewhat less clear, although

6 See Perez, *Invisible Women: Exposing Data Bias in a World Designed for Men*.

7 Comin and Mestieri, “If Technology Has Arrived Everywhere, Why Has Income Diverged.”

8 Bloomberg, “Has the World Managed to Conquer Inflation.”

9 Easterly, “In Search of Reforms for Growth: New Stylized Facts on Policy and Growth Outcomes.”

10 World Bank, “World Development Indicators: Cellular Subscriptions.”

11 Dasgupta, Susmita; Lall, Somik; Wheeler, “Policy Reform, Economic Growth, and the Digital Divide.”

12 Wallsten, “An Econometric Analysis of Telecom Competition, Privatization, and Regulation in Africa and Latin America.”

13 Note that while technical expertise is certainly necessary to achieve the outcome, stabilizing currencies can also be very political—as Zimbabwe and Venezuela have both demonstrated in recent times. And while creating a new vaccine is definitely a technological exercise, actually delivering vaccinations is far more than that—**certainly not without opposition** in countries including Pakistan and Afghanistan, and requiring an extensive institutional structure to deliver. Take telecoms: introducing private provision in other infrastructure sectors has certainly proven anything but purely technical, igniting political confrontation, frequently failing to deliver, sometimes ending in international court cases and collapse. And mobile phone companies face many of the political economy challenges of any private business—capricious enforcement of poorly defined regulations, for example. Blurred lines about both what is technological and what is (efficiently) replicable means that what is amenable to “best practice” or evidence from elsewhere may not always be clear ex ante.

it may have a useful role in delineating what approaches really are universally applicable across countries or identifying which institutional environments better enable such initiatives.

BOTTOM RIGHT? CONTEXTUAL MACRO POLICIES

Industrial policy reform might fall somewhere between top and bottom right: industries are culturally embedded in local communities and established industrial relations but are also connected to exogenous (and somewhat homogenous) regional or global markets and financial institutions. Industries also employ fairly standardised processes and equipment in their production methods. But in the past, economists have sometimes defaulted to national interventions as being technical in nature, attempting to measure and implement “best practice” approaches where in fact local knowledge should have been a key element of the intervention. In that regard, development economics was sometimes dismissed as a separate field on the grounds that policies would work similarly across countries, given that people respond to economic incentives everywhere.¹⁴

As an example, in 1997, two World Bank economists, Craig Burnside and David Dollar, wrote a [paper](#)¹⁵ that looked across countries and suggested aid had a bigger impact on economic growth where recipient countries had “good policies”—as broadly defined by a neoliberal. The paper was influential: it was a factor behind reallocation of aid flows from the World Bank and other aid agencies as well as the creation of the US Millennium Challenge Corporation. But the paper hasn’t stood the test of time: a range of studies have¹⁶ found¹⁷ no link¹⁸ between policies, aid, and growth using similar approaches and more data or varying the analytical approach.

That is not a rare example: some of the most cited papers in development economics using the best methods have turned out to be fragile: change the context or time and the results don’t hold up. There are some things strongly associated with growth, like declining shares of agriculture in output and food in expenditures. But robust and plausibly causal correlates of rising per capita incomes are hard to come by. Investment, education, trade, policies—name your favorite—seem to matter according to regressions using particular data, time periods, countries, and econometric approaches and not matter according to a bunch of others.

It has become a common complaint¹⁹ that cross-country growth regressions are underpinned by some significant assumptions: the “components” of all economies are in some way the same, and that these components of the economy interact with one another in the same kinds of ways, producing economic “laws” or regularities which operate regardless of time or space. In fact, country growth experiences have been extremely heterogeneous, and in a way that is difficult to explain using any one model of economic growth.

Creating a macro policy that fits a specific country context requires technical expertise, but also local knowledge. It may be that the right combination of technical and local knowledge will reliably lead to the “correct” solution in a particular context even though linear relationships between policies and outcomes do not exist: effectively, experts with understanding of control variables will be able to

14 Krueger, “Changing Perspectives on Development Economics and World Bank Research.”

15 Burnside and Dollar, “Aid, Policies, and Growth.”

16 Dalgaard, Carl-Johan; Hansen, “Dalgaard, Carl-Johan; Hansen, Henrik Working Paper On Aid, Growth, and Good Policies.”

17 Easterly, William; Levine, Ross; Roodman, “New Data, New Doubts: A Comment on Burnside and Dollar’s ‘Aid, Policies, and Growth.’”

18 Hansen and Tarp, “Aid and Growth Regressions.”

19 Kenny and Williams, “What Do We Know about Economic Growth? Or, Why Don’t We Know Very Much?”

predict the impact of a particular policy in a particular context. But at the cross-country level, at least, there is insufficient data to conclusively demonstrate that is the case.

The variety in outcomes that arise from similar policy packages is not only because different laws, regulations, and investments enacted the same way can have different impacts in different environments; it is also because the packages are enacted differently; there is frequently a considerable gap between de jure policy and de facto implementation. This has been examined in the context of business regulations,²⁰ structural adjustment reforms,²¹ and regulation of the postal system²² Pritchett (2019) takes from these experiences the lesson that the technical or codifiable knowledge is “at best, a minor constraint on the adoption and effective implementation of targeted programs.” And “what works” is additionally a political question, in that procedures have to be enacted in order to have impact, and that can be prevented by those with power: they may prefer a worse outcome in which they retain relatively more power or resources.

A group of development thinkers who emphasize “navigation by judgment,”²³ “working with the grain,”²⁴ and “problem-driven iterative adaptation”²⁵ in aid programs would suggest a lot of the important activities required to promote development cluster toward the bottom of the figure, with many in the right-hand corner, as would many scholars of successful development experiences like China’s.²⁶ But the fact that much macro policy falls to the bottom right of Figure One does not provide for easy answers. It suggests the importance of local knowledge, but hardly guarantees that such knowledge held by outsiders (or, indeed, locals) will help deliver outcomes.

20 Hallward-Driemeier and Pritchett, “How Business Is Done in the Developing World: Deals versus Rules.”

21 Easterly, “What Did Structural Adjustment Adjust? The Association of Policies and Growth with Repeated IMF and World Bank Adjustment Loans.”

22 Chong et al., “Letter Grading Government Efficiency.”

23 Honig, *Navigation by Judgment: Why and When Top Down Management of Foreign Aid Doesn't Work*.

24 Levy, *Working with the Grain: Integrating Governance and Growth in Development Strategies*.

25 Andrews, Matt; Pritchett, Lant; Woolcook, *Building State Capability: Evidence, Analysis, Action*.

26 Ang, *How China Escaped the Poverty Trap*.

It should be emphasized that cross-country analysis is still a useful tool. Even imperfect answers to important questions are valuable. Not least, cross-country analysis can help generate evidence regarding the plausibility of theoretical predictions. The closer link between learning outcomes and subsequent economic growth²⁷ than between enrolments and subsequent growth,²⁸ for example, does not demonstrate that better learning outcomes will necessarily lead to more rapid development, but it does provide suggestive evidence that schooling absent learning is likely to have limited macro impact. And providing strong evidence that simple growth recipes are not universally applicable was a service in itself.²⁹ Furthermore, there are many important questions for which we cannot do much better than cross-country analysis: not everything can be tested by experiment, including some really important things like interest rate policies.

BOTTOM LEFT? THE EMBEDDED COMPONENTS OF MOST DEVELOPMENT PROJECTS

The left-hand “individual” side of the spectrum contains a number of procedures that are significant to development outcomes. Individual or community procedures can cover activities like deworming (where knowing that the pills work is technological, but programs to deliver them are institutional), issues of police oversight, education interventions, welfare payments, conditional payments to encourage people go to school or take training programs, microcredit, land and business regulation, off-grid utility provision, and local infrastructure construction.

These projects can often be evaluated by a mainstay of development economics journals today: the randomized control trial. A generation of scholars jaded by at best marginally successful efforts to explain the big questions of macroeconomic development have turned instead to (often) small-scale experimentation using carefully designed interventions (frequently) monitored by bespoke surveys. These have an advantage over national-level policies evaluated by cross-country regressions: there is the opportunity and potential data to find out if what worked there, then works here, now.³⁰ Most cross-country regressions lack this identification—we can’t know if what has been uncovered is a plausibly causal relationship from the intervention being tested and the economic growth (or improvements in health or other outcomes) desired. Macroeconomists threw all sorts of statistical innovation at the problem of identification, some more convincing³¹ than others.³² But usually, you

27 Hanushek and Wößmann, “Education and Economic Growth.” and the attached copy is provided by Elsevier for the author’s benefit and for the benefit of the author’s institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues who you know, and providing a copy to your institution’s administrator. All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution’s website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier’s permissions site at: <http://www.elsevier.com/locate/permissionusematerial> Hanushek E A and Wößmann L (2010)

28 Pritchett, “Where Has All the Education Gone?”

29 Despite that service, macro policy universalism has hardly gone away. For example, the World Bank’s Doing Business exercise is still a popular yearly event in which countries get ranked for their laws and regulations: less official steps to get a permit to open a business, more points (it doesn’t matter what the steps are, or if they are followed in practice). And perhaps sometimes, it is even broadly justifiable as we have suggested.

30 With cross-country regressions, economists are limited by (say) 60 years of data and 150 or so countries for a maximum of 9,000 country-years of observations. For good reason people worry about running yearly regressions when looking for potential long-term determinants of growth, so drop that to 1,800 or 900 observations. With hundreds of different potential determinants proposed and hundreds to a power of potential interactions, there’s simply not enough data with which to test macro theories of everything.

31 Galiani et al., “The Effect of Aid on Growth: Evidence from a Quasi-Experiment.”

32 Bazzi and Clemens, “Blunt Instruments: Avoiding Common Pitfalls in Identifying the Causes of Economic Growth.” ISSN: “19457707”, abstract: “Concern has intensified in recent years that many instrumental variables used in

can't run an experiment at the cross-national level—and natural experiments created by (often³³ misguided³⁴) policy choices like cutting off aid funding one side of an arbitrary income threshold are (thankfully) rare.

But every negative result in an RCT (and, overcoming publication bias, more negative results were published if they involved an RCT³⁵) is death to a claim of universal efficacy. Given that, there is some irony that a big complaint about RCTs is similar to the complaint about cross-country regressions 20 years ago: not everywhere is the same. Critics complain that in place of policy universalism we have solution universalism with “proven interventions” based on one or two small-scale RCT results that are taken as “gold standard” evidence of impact and suggested as priorities for global introduction.

So, just as with macroeconomic procedures, it isn't clear that what works in one place at one time on average works as well in another place at another time—or is effective on subjects who are not at the average. For example, returns to investment in agriculture and schooling vary a lot over time in the same place, let alone different places (Rosenzweig and Udry, 2019), and that has implications for RCTs. At the least, this implies many more control variables (and, so, experimental subjects) may be needed to fully understand where the results generated in one place will replicate in others.

Eva Vivalt reports³⁶ considerable heterogeneity across impact evaluations, as well as systematic variation in effect sizes between studies: “smaller studies tended to have larger effect sizes, which we might expect if the smaller studies are better-targeted, are selected to be evaluated when there is a higher a priori expectation they will have a large effect size, or if there is a preference to report larger effect sizes, which smaller studies would obtain more often by chance.”³⁷ Work like Jason Kerwin and Rebecca Thornton's randomized trials of reading interventions in rural Uganda suggest both the considerably different impact on outcomes that can occur with comparatively small design changes to enable scale-up in the same environment as well as the danger of picking the wrong impact measure.³⁸

Even “technological procedures” in health appear to be considerably context dependent with regard to efficacy. The Disease Control Priorities³⁹ exercise has reviewed cost-effectiveness estimates for 93 interventions drawn from 149 studies and, in a few cases, the same intervention in different contexts. These different contexts can suggest significantly different returns to the intervention. DCP reports cost effectiveness estimates that span orders of magnitude for intermittent preventive treatment for

widely-cited growth regressions may be invalid, weak, or both. Attempts to remedy this general problem remain inadequate. We show how a range of published studies can offer more evidence that their results are not spurious. Key steps include: grounding growth regressions in more generalized theoretical models, deployment of new methods for estimating sensitivity to violations of exclusion restrictions, opening the “black box” of GMM with supportive evidence of instrument strength, and utilization of weak-instrument robust tests and estimators. (JEL C52, E23, F35, O41, O47)

33 Kenny, “The Strange and Curious Grip of Country Income Status on Otherwise Smart and Decent People.”

34 Kenny, “The (Sometime) Tyranny of (Somewhat) Arbitrary Income Lines.”

35 Brodeur et al., “Star Wars: The Empirics Strike Back.”

36 Vivalt, “How Much Can We Generalize from Impact Evaluations?”

37 The world is messy and that will be why even the most replicable of ‘graduation’ approaches subject to RCTs (Banerjee et al 2015) tend to suggest rather low economic rates of return, implying they are ameliorative rather than transformative (\$4,545 in investments producing annual returns in the region of \$344, reports Pritchett (2019)).

38 Kerwin and Thornton, “Making the Grade: Understanding What Works for Teaching Literacy in Rural Uganda.” Eva Vivalt notes that the evaluation community has some way to go in reporting all of the details that will allow readers to understand the contextual factors that might have mattered in outcomes. In a survey of papers, she found reported details on who implemented interventions and how they occurred were often sparse and the mechanisms through which the intervention is expected to work were often left unspecified. An effort to record the theories of change suggested in evaluations was abandoned because they were unclear in 90 percent of cases.

39 Horton, “Economic Evaluation Results from Disease Control Priorities, Third Edition.”

malaria amongst infants in Africa, for example. Arnold et al. argue that (even) these reported ranges are often too narrow, with gaps in part reflecting different methodologies and assumptions of underlying studies but also practical issues of health system capacity. They note that the cost effectiveness of caesarean sections reported by Disease Control Priorities for lower-middle-income countries is between \$1,600 to \$2,600 US\$ per DALY averted, but that results based on underlying modelling estimates suggest a range from \$251 to \$3,462.⁴⁰ Different worms and different worm loads in different places also account for some of the variation in the reported efficacy⁴¹ of mass drug administration/chemotherapy in the literature surrounding the “worm wars,” different institutional capacities and design choices may account for some of the rest.⁴²

Of course, we still have the methods to make quantitative comparisons and none of this suggests otherwise. For example, we can measure that vector control for Dengue fever is likely to be an expensive intervention compared to preventive chemotherapy for onchocerciasis (perhaps yielding less than one hundredth of the DALYs per dollar spent). But it does suggest concerns with global best practice or international validity occur even in individual comparatively technological areas like health.

RCT advocates recognize and address the generalizability puzzle,⁴³ asking questions like “what’s the theory behind the program and do local conditions hold?”, “how strong is the evidence for required behavioral change and that implementation can be carried out well?” An increased recognition of the importance of context to validity might be suggested by the greater geographic specificity of paper titles reporting small area studies and growing concern with control variables (do children given deworming pills wear shoes, for example).

To quote Abhijit Banerjee and Ester Duflo: “If we were prepared to carry out enough experiments in varied enough locations, we could learn as much as we want to know about the distribution of the treatment effects.” And RCTs have been used that way: Sandefur and colleagues scaled up an experiment in a small region of Kenya that had successfully used contract teachers to improve scores only to find it didn’t work⁴⁴ when run by the Kenyan government, which ended up regularizing many of the teachers involved. The work⁴⁵ of RTI’s Ben Piper and colleagues with the Kenyan Ministry of Education found that a successful literacy intervention did replicate at the national level. Again, a number of interventions including the recently cancelled *Oportunidades* program in Mexico were tested with RCTs at national scale. There is also mounting randomized evidence of the impact⁴⁶ of cash transfers

40 Those underlying estimates ignore benefits to the neonatal (despite the fact the caesarean usually prevents neonatal death: Arnold et al., “A One Stop Shop for Cost-Effectiveness Evidence? Recommendations for Improving Disease Control Priorities.”)

41 Majid, Kang, and Hotez, “Resolving ‘Worm Wars’: An Extended Comparison Review of Findings from Key Economics and Epidemiological Studies.”

42 Lant Pritchett and Justin Sandefur [examine](#) randomized control trial evidence covering the effects of class size and private schooling on outcomes to demonstrate the limits to assuming what is valid evidence in one place carries across into policy implications for another. They suggest the estimating impact from non-randomized outcomes from public versus private schools or bigger versus smaller class size in your context is a more reliable guide to the actual impact (as measured by an RCT in your context) than is applying the results of an RCT from somewhere else. To put it another way, the effect on learning outcomes driven by the fact that students in public schools have different characteristics than students in private schools (which is the reason to randomize) is smaller than the effect on learning outcomes driven by the gap between public and private school performance (accounting for student characteristics) across regions. In this case, RCTs from elsewhere are worse than observations of averages from here in predicting the impact of a policy change.

43 Bates, Mary Ann; Glennerster, “The Generalizability Puzzle.”

44 Bold et al., “Scaling Up What Works: Experimental Evidence on External Validity in Kenyan Education Scaling Up What Works: Experimental Evidence on External Validity in Kenyan Education We Are Indebted to the Staff of the Ministry of Education, the National Examination.”

45 Piper et al., “Scaling up Successfully: Lessons from Kenya’s Tusome National Literacy Program.”

46 Ralston, Andrews, and Hsiao, “The Impacts of Safety Nets in Africa What Are We Learning ?”

(and conditional cash transfers) on outcomes in a range of different settings—even if that impact still varies a lot and the evidence is mixed that impact is sustained once the transfers stop.

But despite this recognition, it is safe to say most policy questions haven't been subject to randomization anywhere, let alone in enough places to point towards a fairly robust (or not) relationship regardless of context.⁴⁷ This all suggests that to get the most out of RCT approaches where they can be used, they will have to become a lot more widespread than they are now. To quote Pritchett and Sandefur,⁴⁸ the importance of local context “is not an argument against randomization as a methodological tool for empirical investigation and evaluation. It is actually an argument for orders of magnitude more use of randomization.”⁴⁹

The contribution of hundreds of academics running RCTs, including Banerjee and Duflo, has positively impacted millions of lives, demonstrated what is possible in terms of rigorous analysis of policy, and created a large cadre of people who can run (or at least interpret) experiments. But given the importance of local context to many of the types of procedures RCTs have been used to test, RCTs themselves will not be able to be scaled to the extent needed if the model has to involve outsiders with the awesome talents and skill sets of a Duflo or Banerjee, especially if those outsiders are primarily motivated by the need to publish in journals.⁵⁰ The approach of bespoke one- or few-off experiments in conducive environments involving specially commissioned experiments and surveys is similarly not very scalable.

One approach to help deal with this problem is larger RCTs with more arms occurring in more locations (the approach adopted⁵¹ in the JPAL integrated rural livelihoods project). A linked approach would be running bigger experiments over larger populations using regularly collected administrative data employing government staff rather than outside experts. But the sad fact is that there aren't lots of governments in developing countries running or even commissioning their own RCTs. It is fairly rare even in rich countries. The UK Government What Works initiative suggests it has commissioned or supported over 160 trials over five years⁵²—or an average of only about 30 a year. And the What Works Network suggests it has financed “over 10 percent of all robust education trials in the world”—the UK accounts for a little under 0.6 percent of the world's population aged 0-14, suggesting Britain is far better than most.

47 This links to the debate over the relative benefits provided by the comparatively certain knowledge provided by RCTs of micro interventions in specific locations as compared to the considerably weaker knowledge provided by analysis of macroeconomic policy impact using cross-country or anecdotal approaches, to which the answer is likely to be “it depends.”

48 Pritchett and Sandefur, “Context Matters for Size: Why External Validity Claims and Development Practice Don't Mix.”

49 Interestingly, Banerjee and Duflo have argued that the most valuable contribution of RCTs is through the close collaboration between researchers and implementers, which allows for iteration of designs, creative experimentation in which the two together “think out of the box and learn from successes and failures” as part of a long-term relationship. Duflo has suggested the model of “plumbers” who “try to predict as well as possible what may work in the real world, mindful that tinkering and adjusting will be necessary since our models give us very little theoretical guidance on what (and how) details will matter.”

50 To illustrate that problem, take a line from Banerjee and Duflo's review: “the experimental work in the mid-1990s (e.g., Glewwe et al. 2004, 2009; Banerjee et. al. 2005) was aimed at answering very basic questions...” Despite its focus on basic questions, that work carried out in the mid-1990s only got published in the mid-aughts. People motivated by the desire to publish in top economics journals will spend a lot of time looking for experiments that can be published in top economics journals, thinking about innovative approaches that are top-journal worthy and spend years going through submissions, revisions and resubmissions to get those papers into top journals. This may be a good way to run the academic discipline of economics, it is not necessarily the most efficient use of time and resources from the point of view of learning more about what works in development.

51 Banerjee et al., “A Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries.”

52 Cabinet Office, “The What Works Network: Five Years On.”

Donors should support mainstreaming of evidence-based policy including the data to support it through bigger, better administrative data. And, more broadly, for the development community, it should not be about looking for places and programs where a particular evaluation approach amenable to journal publication can be tried, but instead looking for big important questions and then trying to find the best way to answer them: focus on procedures that could potentially really matter at scale, then test them using the most rigorous available method. That can be RCTs under some circumstances (interventions that are done at the individual or local level, expected to have a reasonably short-term impact with easily measurable—and measured—results) or other techniques for interventions like industrial policy.

The framework with examples is shown in Figure 2.

Figure 2: What Interventions Fit Where?

		Scale of Intervention (spectrum)	
		Individual	(Inter)national
Nature of Intervention (spectrum)	Technical	Vaccinations	Monetary Policy Fiscal Policy Mobile Telecomms deregulation
	Cultural	Deworming Hygiene Education	Industrial Policy Education System Reform

APPLYING THE FRAMEWORK: LESSONS FOR DONORS

To summarize what this all might imply using the four corners of Figure 1:

- **Top Left:** If the process is technological/individual/local—including new technologies like vaccines, solar cells, or off-grid toilets, for example—there is a large role for laboratory research, which should be conducted wherever is most efficient. But note that even if inventions are technical, their adoption may be highly dependent on cultural factors, thus also demanding local research. And the focus of efforts to increase the pipeline of inventions will require considerable thinking about where limited aid resources can have the greatest impact.

- **Top Right:** If the process is “technological”/national, there is some role for formal political economy research even without hope of definitive answers—focusing on topics like monetary policy, industrial and trade policy, environmental and business regulation, the organization of ministerial functions, the impact of subsidy reform, and so on. This would combine approaches from multiple disciplines but also multiple perspectives. Related to that, there is a large role to support the build-up of experience and knowledge of “best practice” and the conditions necessary for it to work. This would look more like “technical assistance” than research and development.
- **Bottom Left:** If the process is embedded/individual, there is a significant role for research, but it has to be localized, with implications for how it can be done sustainably. Monitored experimentation by implementing institutions potentially supported with donor initiatives in global data collection and technical support is likely to be more scalable than direct financing and implementation of experimental pilots and bespoke surveys.
- **Bottom Right:** If the process is embedded/national, neither formal research nor international experience and best practice will have a major role to play in developing solutions. Perhaps outsiders can play supportive roles as honest brokers and conveners. Infant industry protection might provide a good example.

SPENDING AID MONEY ON RESEARCH FOR INVENTIONS

It is worth stepping back to point out that the generation and spread of knowledge is a central, indispensable element of development. Different returns to a given stock of labour and capital—total factor productivity—accounts for most of the global variation in income worldwide. More than levels of investment, it is how an investment is used that determines wealth and poverty. And technological advance (broadly defined) accounts for higher total factor productivity. Again, the invention and spread of technologies such as oral rehydration, bednets, vaccines and antibiotics means life expectancies in the world’s poorest countries are considerably higher than those in far richer countries in the past.⁵³

Aid resources have played a vital part in developing and rolling out technologies and innovations that have had a significant impact on productivity and the quality of life—from Norman Borlaug’s work on the dwarf and resistant plant varieties of the green revolution supported by the Rockefeller Foundation through the advance market commitment for the pneumococcal vaccine supported by Gavi to DFID’s backing for the development of M-PESA mobile finance in Kenya. The economic returns to such advances can be huge (DFID’s M-Pesa investment was only one million pounds, annual mobile payments in the country are now worth over 50 percent of Kenya’s GDP).⁵⁴

53 Kenny Getting Better

54 BBC, “Money via Mobile: The M-Pesa Revolution.”

But it is also worth noting the likely limits to aid-funded technology research. Barder and Kenny (2015) estimate developing countries spent about \$300 billion on R&D, and neglected tropical diseases alone see about \$3.2 billion in R&D. Whatever UK aid does will still be a drop in the bucket compared to such totals (current total UK R&D ODA spending is about GBP 747 million, or \$934 million). To take a specific example, it is surely true that if the world developed a solar film technology that dramatically reduced the cost of solar power, that would have a dramatic impact on global development.⁵⁵ It is also true that solar power research is already the subject of an considerable global effort, with about \$10 billion spent on solar R&D in 2018⁵⁶ --give or take, ten times the total amount that the UK spends on ODA for R&D.⁵⁷

There is still a considerable role for ODA-funded R&D not least because DFID's primary goal in research spending is social value for developing countries as a whole rather than private or narrow national good. DFID can leverage private spending through mechanisms including pull incentives, and the department can concentrate on demonstration effects, crowding in other developed country donors to joint projects. But the scale issue clarifies why the UK should focus innovation research efforts in places and on activities that are likely to be a big public good in terms of improving the lives of the world's poorest people. This is likely to include technologies that reduce the need for scarce governance capacity and capital—for example the off-grid toilet being backed by the Gates Foundation that removes the need for expensive and institutionally complex piped sewage systems, or research into cheap, easily applied and long-lasting road surface treatments that survive in tropical climates. It will also include seed varieties suitable for the climate and capital constraints of smallholder farming in developing countries, cheap prophylactics for disease conditions concentrated in tropical climates, and genetically modified mosquitoes.

Cutting edge biomedical or road surface advances are likely to involve researchers from the “Global North.” But this does not necessarily imply UK researchers nor does the location of research have to be determined prior to the subject of research being decided. There is a considerable role for ‘pull’ mechanisms, including advance market commitments that direct research effort to answering a development challenge specific enough to be legally contractable (delivery of a pneumococcal vaccine with a formulation covering at least 60 percent of the invasive disease isolates in the target region, including serotypes 1, 5 and 14, designed to prevent disease among children with restricted contra-indications, scheduling compatible with immunization programs –and so on).⁵⁸ And given context matters even for the efficacy of inventions, new technological solutions should be developed and evaluated preferably by local research capacity in developing countries potentially with the support of outside researchers.

SPENDING MONEY ON POLICY RESEARCH

Aid-financed research on policy innovation can significantly impact development outcomes. We have seen that RCTs have played a role in improving program design across a range of projects, for example. And there is now experimental evidence that, at least in the setting of Brazilian municipalities, policy makers respond to research on policy reform by changing policies.⁵⁹

Nonetheless UK R&D spend on policy innovation will also have to be focused if it is to achieve more

55 Guardian, “Rory Stewart: Boris Johnson Win Would Bring DfID Tenure to ‘heartbreaking’ End.”

56 Joint Research Centre. European Commission, “PV Status Report 2018.”

57 Ritchie, “ODA for Research & Development: Too Much of a Good Thing?”

58 GAVI, “AMC Legal Agreements.”

59 Hjort et al., “How Research Affects Policy: Experimental Evidence from 2,150 Brazilian Municipalities.”

than a marginal impact—it is simply insufficient to support the level of experimentation necessary given the (apparent) limited portability of many policy models. Should the UK be funding (as a matter of priority) evaluations of programs that are *unlikely* to be easily replicated at scale or in other regions or countries (in cases where this can be predicted a priori)? No, because outside bespoke evaluations are unlikely to be a sustainable or affordable model. And for all of the concerns about capacity, African institutions already have research advantages over Western researchers in terms of local knowledge and networks as well as low costs.

UK ODA should support the development of local capacity to solve challenges that (because of their embedded/individual or embedded/national characteristics) will have local solutions that may not be portable. This may include support for training and skills development in evaluation delivered either locally or elsewhere, challenge funding to finance evaluations of particular local solutions to be led by local institutions, and global support through (for example) satellite solutions that provide actionable data at the local level. When context matters, the aim should be to empower locals to ask the right questions figure out answers (not to propose an answer and test it locally). When context doesn't matter, the aim should be to do research where it is most efficiently done or, where there are multiple options for doing it efficiently, where the most positive spillovers of research activity and expenditure can be had.

In that regard, it is worth noting the very limited global aid support for local research in developing countries. Sub-Saharan Africa outside of South Africa has 56 scientists per million inhabitants (compared to 4,181 per million in the UK)⁶⁰ but the system of external financing and enforced “partnering” in aid spending in this area means that research priorities even for many of these scientists are set outside of the continent, where most of the money is also spent. In 2012, 79 percent of research in Southern Africa involved collaboration with external partners, for example.⁶¹ The existing mechanisms for strengthening African research institutions directly are extremely small. Alex Ezeh and Jessie Lu compile a list of six ongoing initiatives to build research capacity in sub-Saharan Africa financed by donors including the World Bank, DFID, the US National Institutes of Health and a number of foundations which have a *combined* annual budget of about \$100 million a year (or around \$2 million per country).⁶²

From Local to Global

Because there is a global role to evaluate whether if global or at least regional “best practices” are in fact emerging from local research, and because there are intermediate procedures in the middle of the spectrum, there are roles for intermediate approaches. It may be that the broad and significant determinants of cost effectiveness in health procedures in different contexts are few enough that a global effort to create country-specific estimates would be far more efficient than individual countries attempting to create the whole evidence base for bespoke country health technology assessments to determine the most efficient way to spend health funding, for example.⁶³

Expanding the capacity (and budget) of the DFID-funded International Decision Support Initiative could allow it to further examine what lies behind the variance in cost effectiveness of interventions across countries –not least whether it is primarily different assumptions or methodologies as opposed

60 UNESCO, “UNESCO Science Report 2010.”

61 Ezeh and Lu, “Transforming the Institutional Landscape in Sub-Saharan Africa: Considerations for Leveraging Africa’s Research Capacity to Achieve Socioeconomic Development.”

62 Ibid

63 See Chalkidou et al., “We Need a NICE for Global Development Spending.”

to different contexts. Depending on that answer, and building on prevalence and demographic data, it might be able to provide standardized cost effectiveness estimates based on prevalence and population, alongside economic factors. This might provide baseline data for national health effectiveness exercises. Something similar might apply to education interventions, local government reform, welfare and micro-finance/ small scale employment projects. Even if the answer developed by such reviews is that local conditions dominate best practice design in determining outcomes, the lack of portability of approaches is a useful result for policy makers.

Such exercises should be conducted wherever they are most efficiently done. But given that they are not original research endeavours, at the least they could involve a global consortium of research institutions that combined technical expertise as well as local knowledge.

Finally, challenges that are national will lie along a spectrum from embedded to “technological,” with the end that is closer to technological holding more promise for “best practice” and formal research into solutions and mechanisms. This work is a global public good and has indeed been a priority of international organizations, including the UNDP and the World Bank,⁶⁴ although with a mixed track record of success.

HOW THE UK IS SPENDING R&D ODA

A considerable portion of recently expanded UK R&D financing is spent on worthy research around high-priority issues including health challenges specific to tropical countries—for example the Ross Fund, which supports research into new products for infectious and tropical diseases alongside implementation programs for malaria and neglected tropical diseases. Support also covers approaches to speed economic growth through the demand-led International Growth Center. However, the spending is overwhelmingly bilateral, UK based,⁶⁵ and a significant portion is allocated based on demand for funding from UK researchers rather than the urgency and tractability of the research topic in developing countries.

Speaking to the lack of focus, 45 percent of R&D ODA is classified for “research/scientific institutions”—an “other” category for sectoral allocation. Sixty-two percent of R&D ODA is allocated to ‘developing countries, unspecified’, compared to 25 percent allocated to individual countries and 13 percent to regions. This will reflect in part the fact that about half of UK R&D ODA is spent through the Department of Business, Energy and Industrial Strategy on projects housed in UK universities many of which have been reclassified as aid.

The GBP1.5 billion Global Challenges Research Fund to be spent between 2016-21 supports UK institutions to address “the problems faced by developing countries, whilst developing our ability to deliver cutting-edge research.”⁶⁶ The starting point for research and innovation funded through GCRF, suggests the strategy, should be “a significant problem or development challenge.” The specific (which is to say broad) challenge areas outlined by the GCRF are Equitable Access to Sustainable Development, Sustainable Economies and Societies and Human Rights, Good Governance and Social Justice.

64 Banerjee, Abhijit; Deaton, Angus; Lustig, Nora; Rogoff, “An Evaluation of World Bank Research, 1998–2005.”

65 Some GCRF projects have been awarded to Southern institutions and the extent to which partners in the Global South have played a leading role in problem identification and the design, definition, and development of the proposed approach forms an important part of the criteria for funding decisions according to the strategy. But, in line with the aim to “develop our ability,” the considerable majority of the funding appears to have gone to UK institutions.

66 BEIS, “UK Strategy for the Global Challenges Research Fund (GCRF).”

Research Councils, Academies and the UK Space Agency, collective recipients of four fifths of the allocated funding reported in the 2015 spending review, use thematic calls for applications followed by a selection procedure based on scientific and academic merit. UK funding Councils, recipients of 20 percent of allocated funding, simply pass on resources according to their standard allocation formula.⁶⁷

The UK academic incentive structure is not (and should not be) designed to maximize utility of research for development outcomes. But academics learning and writing about developing countries for the sake of “social science smacking of orientalism” is not a global public good.⁶⁸ The 2017 ICAI review of the GCRF suggested that adherence to the Haldane Principle that scientists should decide on the merits of research proposals is probably reasonable for basic research, but UK aid should be targeted specifically at unique challenges of developing countries, which, rather than UK based technical researchers, are more readily identified by development practitioners, policymakers, and, above all, the residents of these countries themselves: their policy experts, their leaders, and their CSOs. This is surely correct. The ICAI review also suggested a more explicit prioritization of capacity building in southern institutions, again this seems key, along with increased support for multilateral solutions.

UK AID POLICY REFORM FOR IMPACT

For new technologies, UK ODA should be prioritized on innovations to tackle public good problems specific to the world’s poorest countries, such as cheap sanitary solutions and prophylactics for tropical diseases. The Secretary of State for International Development should consider setting up a commission made up of representatives of developing countries, development and scientific experts to draw up a list of potential innovations that would (i) ameliorate or solve public policy challenges specific to developing countries and, (ii) would require comparatively small additional research and development steps to bring to market. The commission would also propose appropriate mechanisms to incentivize the research. In cases where the likely solution is understood well enough that it can be contracted, the approach might include prizes or advance market commitments. In other cases, the approach might include push support through institutions, preferably selected competitively using evaluation criteria developed by the commission. Another approach the commission might back where appropriate would involve buying out patent owners of existing technologies.

For national-level policy procedures and knowledge creation in areas where local information may at least be informative for cross-country decision-making (including in areas such as effectiveness in health spending), UK ODA could provide greater support to initiatives including the International Growth Center and the International Decision Support Initiative, perhaps including a stronger commitment to demand-led research and capacity building in developing countries and a greater willingness to competitively select host institutions using a global competition.

For local policy effectiveness research, UK ODA should focus on interventions to support local capacity including (i) embedding researchers with governments to provide technical support and capacity building, (ii) providing funding to local universities and research institutions to answer policy research requests, and (iii) supporting data collection efforts perhaps in particular through global approaches including satellite data systems.

⁶⁷ The Independent Commission for Aid Impact, “Global Challenges Research Fund - A Rapid Review.”

⁶⁸ I thank Justin Sandefur for the term

REFERENCES

- Andrews, Matt; Pritchett, Lant; Woolcock, Michael. *Building State Capability: Evidence, Analysis, Action*. Oxford University Press, 2017. <https://bsc.cid.harvard.edu/building-state-capability-evidence-analysis-action>.
- Ang, Yuen Yuen. *How China Escaped the Poverty Trap*. Cornell University Press, 2016.
- Arnold, Matthias, Susan Griffin, Jessica Ochalek, Paul Reville, and Simon Walker. "A One Stop Shop for Cost-Effectiveness Evidence? Recommendations for Improving Disease Control Priorities." *Cost Effectiveness and Resource Allocation* 17, no. 1 (2019): 7. <https://doi.org/10.1186/s12962-019-0175-6>.
- Banerjee, Abhijit; Deaton, Angus; Lustig, Nora; Rogoff, Ken. "An Evaluation of World Bank Research, 1998–2005," 2006.
- Banerjee, Abhijit, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert, and Christopher Udry. "A Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries." *Science* 348, no. 6236 (2015). <https://doi.org/10.1126/science.1260799>.
- Bates, Mary Ann; Glennerster, Rachel. "The Generalizability Puzzle." *Stanford Social Innovation Review*, 2017. https://ssir.org/articles/entry/the_generalizability_puzzle.
- Bazzi, Samuel, and Michael A. Clemens. "Blunt Instruments: Avoiding Common Pitfalls in Identifying the Causes of Economic Growth." *American Economic Journal: Macroeconomics* 5, no. 2 (2013): 152–86. <https://doi.org/10.1257/mac.5.2.152>.
- BBC. "Money via Mobile: The M-Pesa Revolution," 2017. <https://www.bbc.co.uk/news/business-38667475>.
- BEIS. "UK Strategy for the Global Challenges Research Fund (GCRF)," 2017. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/623825/global-challenges-research-fund-gcrf-strategy.pdf.
- Bloomberg. "Has the World Managed to Conquer Inflation," 2019. <https://www.bloomberg.com/news/articles/2013-01-28/has-the-world-managed-to-conquer-inflation>.
- Bold, Tessa, Mwangi Kimenyi, Germano Mwabu, Alice Ng, Justin Sandefur, Tessa Bold Mwangi Kimenyi Germano Mwabu Alice Ng, Salome Ong, et al. "Scaling Up What Works: Experimental Evidence on External Validity in Kenyan Education Scaling Up What Works: Experimental Evidence on External Validity in Kenyan Education We Are Indebted to the Staff of the Ministry of Education, the National Examination," 2013. www.cgdev.org/http://www.cgdev.org/.
- Brodeur, Abel, Lé Mathias, Sangnier Marc, and Zylberberg Yanos. "Star Wars : The Empirics Strike Back." *Institute for the Study of Labor*, 2013. <https://doi.org/10.1257/app.20150044>.
- Burnside, Craig, and David Dollar. "Aid, Policies, and Growth." *American Economic Review* 90, no. 4 (2000): 847–68. <https://doi.org/10.1257/aer.90.4.847>.
- Cabinet Office. "The What Works Network: Five Years On," no. January (2018): 1–37. https://what-works.blog.gov.uk/wp-content/uploads/sites/80/2018/04/6.4154_What_works_report_Final.pdf

f%0Ahttps://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/677478/6.4154_What_works_report_Final.pdf.

Chalkidou, K, A J Culyer, A Glassman, and R Li. “We Need a NICE for Global Development Spending.” *F1000Research* 6, no. 1223 (2017). <https://doi.org/10.12688/f1000research.11863.1>.

Chong, Alberto, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. “Letter Grading Government Efficiency.” *Journal of the European Economic Association* 12, no. 2 (2014): 277–99. <https://doi.org/10.1111/jeea.12076>.

Comin, Diego, and Martí Mestieri. “If Technology Has Arrived Everywhere, Why Has Income Diverged.” *American Economic Journal: Macroeconomics* 10, no. 3 (2018): 137–78.

Dalgaard, Carl-Johan; Hansen, Henrik. “Dalgaard, Carl-Johan; Hansen, Henrik Working Paper On Aid, Growth, and Good Policies.” *CREDIT Research Paper* 00/17 (2000). <https://www.econstor.eu/bitstream/10419/81761/1/00-17.pdf>.

Dasgupta, Susmita; Lall, Somik; Wheeler, DAvid. “Policy Reform, Economic Growth, and the Digital Divide,” 2001. <http://documents.worldbank.org/curated/en/935091468741308186/pdf/multi0page.pdf>.

Easterly, William; Levine, Ross; Roodman, David. “New Data, New Doubts: A Comment on Burnside and Dollar’s ‘Aid, Policies, and Growth,’” 2000. [https://doi.org/10.1016/0021-9045\(89\)90093-2](https://doi.org/10.1016/0021-9045(89)90093-2).

Easterly, William. “In Search of Reforms for Growth: New Stylized Facts on Policy and Growth Outcomes,” 2018. [https://www.gc.cuny.edu/CUNY_GC/media/CUNY-Graduate-Center/PDF/Programs/Economics/Other docs/Easterly-Policy-Reforms-and-Growth.pdf](https://www.gc.cuny.edu/CUNY_GC/media/CUNY-Graduate-Center/PDF/Programs/Economics/Other%20docs/Easterly-Policy-Reforms-and-Growth.pdf).

———. “What Did Structural Adjustment Adjust? The Association of Policies and Growth with Repeated IMF and World Bank Adjustment Loans,” 2002. <https://doi.org/10.1016/j.jdevco.2003.11.005>.

Ezeh, Alex, and Jessie Lu. “Transforming the Institutional Landscape in Sub-Saharan Africa: Considerations for Leveraging Africa’s Research Capacity to Achieve Socioeconomic Development,” no. July 2019 (2019). www.cgdev.orgwww.cgdev.org.

Galiani, Sebastian, Stephen Knack, Lixin Colin Xu, and Ben Zou. “The Effect of Aid on Growth: Evidence from a Quasi-Experiment.” *NBER Working Paper*, 2016. <https://doi.org/10.1007/s10887-016-9137-4>.

GAVI. “AMC Legal Agreements,” 2019. <https://www.gavi.org/investing/innovative-financing/pneumococcal-amc/amc-legal-agreements/>.

Guardian. “Rory Stewart: Boris Johnson Win Would Bring Dfid Tenure to ‘heartbreaking’ End,” 2019. <https://www.theguardian.com/global-development/2019/jun/26/rory-stewart-boris-johnson-win-would-bring-dfid-tenure-to-heartbreaking-end>.

Hallward-Driemeier, Mary, and Lant Pritchett. “How Business Is Done in the Developing World: Deals versus Rules.” *Journal of Economic Perspectives* 29, no. 3 (2015): 121–40. <https://doi.org/10.1257/jep.29.3.121>.

Hansen, Henrik, and Finn Tarp. “Aid and Growth Regressions.” *Journal of Development Economics* 64, no.

2 (April 1, 2001): 547–70. [https://doi.org/10.1016/S0304-3878\(00\)00150-4](https://doi.org/10.1016/S0304-3878(00)00150-4).

Hanushek, E A, and L Wößmann. “Education and Economic Growth.” In *International Encyclopedia of Education*, edited by Barry Peterson, Penelope ; Baker, Eva; McGaw, 2:245–52. Oxford: Elsevier, 2010.

Hjort, Jonas, Diana Moreira, Gautam Rao, and Juan Francisco Santini. “How Research Affects Policy: Experimental Evidence from 2,150 Brazilian Municipalities.” Working Paper Series, June 2019. <https://doi.org/10.3386/w25941>.

Honig, Dan. *Navigation by Judgment: Why and When Top Down Management of Foreign Aid Doesn't Work*. Oxford University Press, 2018.

Horton, Susan. “Economic Evaluation Results from Disease Control Priorities, Third Edition.” *Disease Control Priorities, Third Edition*, 2018. <https://www.ncbi.nlm.nih.gov/books/NBK525294/>.

Joint Research Centre. European Commission. “PV Status Report 2018,” 2018. <https://doi.org/10.2760/826496>.

Kenny, C., and D. Williams. “What Do We Know about Economic Growth? Or, Why Don't We Know Very Much?” *World Development* 29, no. 1 (2001): 1–22. [https://doi.org/10.1016/S0305-750X\(00\)00088-7](https://doi.org/10.1016/S0305-750X(00)00088-7).

Kenny, Charles. “The (Sometime) Tyranny of (Somewhat) Arbitrary Income Lines,” 2017. <https://www.cgdev.org/blog/sometime-tyranny-somewhat-arbitrary-income-lines>.

———. “The Strange and Curious Grip of Country Income Status on Otherwise Smart and Decent People,” 2014. <https://www.cgdev.org/blog/strange-and-curious-grip-country-income-status-otherwise-smart-and-decent-people>.

Kerwin, Jason T, and Rebecca Thornton. “Making the Grade: Understanding What Works for Teaching Literacy in Rural Uganda,” 2015. <https://pdfs.semanticscholar.org/0cdb/9ac076286f1d0c836fa7c99c14813cf8736a.pdf>.

Krueger, Anne O. “Changing Perspectives on Development Economics and World Bank Research.” *Development Policy Review* 4, no. 3 (1986): 195–210. <https://doi.org/10.1111/j.1467-7679.1986.tb00381.x>.

Levy, Brian. *Working with the Grain: Integrating Governance and Growth in Development Strategies*. Oxford University Press, 2014.

Majid, Muhammad Farhan, Su Jin Kang, and Peter J. Hotez. “Resolving ‘Worm Wars’: An Extended Comparison Review of Findings from Key Economics and Epidemiological Studies.” *PLOS Neglected Tropical Diseases* 13, no. 3 (2019). <https://doi.org/10.1371/journal.pntd.0006940>.

Perez, Caroline Criado. *Invisible Women: Exposing Data Bias in a World Designed for Men*. Vintage Digital, 2019.

Piper, Benjamin, Joseph Destefano, Esther M. Kinyanjui, and Salome Ong’ele. “Scaling up Successfully: Lessons from Kenya’s Tusome National Literacy Program.” *Journal of Educational Change* 19, no. 3 (2018): 293–321. <https://doi.org/10.1007/s10833-018-9325-4>.

Pritchett, Lant. “Where Has All the Education Gone?” *World Bank Economic Review* 15, no. 3 (2001): 367–91.

Pritchett, Lant, and Justin Sandefur. "Context Matters for Size: Why External Validity Claims and Development Practice Don't Mix," 2013. <https://doi.org/10.2139/ssrn.2364580>.

Ralston, Laura, Colin Andrews, and Allan Hsiao. "The Impacts of Safety Nets in Africa What Are We Learning?," 2017. <https://doi.org/10.1596/1813-9450-8255>.

Ritchie, Euan. "ODA for Research & Development: Too Much of a Good Thing?," 2019. <https://www.cgdev.org/blog/oda-research-development-too-much-good-thing>.

Robinson, Lee, Euan Ritchie, and Charles Kenny. "UK Research Aid: Tied, Opaque, and Off-Topic?," 2019. <https://www.cgdev.org/sites/default/files/uk-research-aid-tied-opaque-and-topic.pdf>.

The Independent Commission for Aid Impact. "Global Challenges Research Fund - A Rapid Review," no. September (2017): 40. <https://icai.independent.gov.uk/wp-content/uploads/ICAI-GCRF-Review.pdf>.

UNESCO. "UNESCO Science Report 2010," 2010. <https://unesdoc.unesco.org/ark:/48223/pf0000189883>.

Vivalt, Eva. "How Much Can We Generalize from Impact Evaluations?" *Working Paper, Australian National University*, 2019.

Wallsten, Scott J. "An Econometric Analysis of Telecom Competition, Privatization, and Regulation in Africa and Latin America." *The Journal of Industrial Economics* 49, no. 1 (2003): 1–19. <https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-6451.00135>.

World Bank. "World Development Indicators: Cellular Subscriptions," 2019. <https://data.worldbank.org/indicator/IT.CEL.SETS.P2?locations=ET-TZ&page=4>.

———. "World Development Indicators: Immunization," 2019. <https://data.worldbank.org/indicator/SH.IMM.IDPT>.

———. "World Development Indicators: Under 5 Mortality," 2019. <https://data.worldbank.org/indicator/SH.DYN.MORT?locations=AF>.



WWW.CGDEV.ORG

This work is made available under the terms of the Creative Commons Attribution-NonCommercial 4.0 license.

CHARLES KENNY is a senior fellow at the Center for Global Development.

EUAN RITCHIE is a research associate at the Center for Global Development.

LEE ROBINSON is a research associate at the Center for Global Development.