The Need for Impact Evaluation in Electricity Access Research

Patrick Bayer[a,1,*], Ryan Kennedy[b,1], Joonseok Yang[c,1], Johannes Urpelainen[c,1]

[a] *University of Strathclyde, School of Government & Public Policy, McCance Building, 16 Richmond Street, Glasgow, G1 1QX, Scotland, UK*
[b] *University of Houston, Department of Political Science, Philip Guthrie Hoffman Hall, 3551 Cullen Boulevard, Houston, TX 77204, USA*
[c] *Columbia University, Department of Political Science, IAB, 420 W. 118th Street, New York City, NY 10027 USA*
[d] *Johns Hopkins University, School of Advanced International Studies, 1619 Massachusetts Avenue, NW, Washington, DC 20036, USA*

—Forthcoming in Energy Policy—

Abstract

Universal household electrification is a key component of the United Nations Sustainable Development Goals, but the evidence base for social and economic impacts of electricity access remains unclear. Here we report results from a systematic review of impact evaluations of household electrification based on five key outcome measures. We only find 31 studies that conduct statistical hypothesis tests to assess impacts. Among these, seven draw on a randomized experiment designed for causal inference. The randomized experimental studies generate fewer positive results than observational or quasi-experimental studies, such as correlational, instrumental variable, and difference-in-differences designs. These results call for a reassessment of what we know about the impacts of household electrification. They also call for major investment in impact evaluation of electricity access using randomized controlled trials, with a particular focus on when and how energy access interventions can furnish large benefits to their intended beneficiaries. Large-scale impact evaluations using experimental methods will require close collaboration between policymakers and researchers.

Keywords: impact evaluation; electricity access; observational and experimental methods; causal inference; sustainable development.

[*] Corresponding author. Replication data available from https://doi.org/10.7910/DVN/LVPPP7.

Email addresses: patrick.bayer@strath.ac.uk (Patrick Bayer), rkennedy@uh.edu (Ryan Kennedy), jy2385@columbia.edu (Joonseok Yang), johannes.u@jhu.edu (Johannes Urpelainen)

1. Introduction

More than one billion people still lack access to electricity at home (IEA, 2017). The United Nations has declared energy access as one of the Sustainable Development Goals (SDG 7) and estimates that an annual investment of between US$1,058-1,266 billion is required to meet the goal by 2030 (UNDP and UN Environment, 2018). The rationale for investments in household electrification is that it could contribute to social and economic development (SE4ALL, 2017) and improve quality of life (Lambert et al., 2014). The evidence base for these expected impacts, however, is unclear. Some recent studies have suggested that the impact of household electrification may be minimal for poor households and that there is a significant opportunity cost to spending heavily on household electrification instead of other infrastructure and energy investments (Lee, Miguel, and Wolfram, 2018; Lenz et al., 2017; Burlig and Preonas, 2016). Due to the lack of a systematic review of the empirical literature, policymakers lack an understanding of the development impact of household electrification and where the gaps in our knowledge remain.

Others have identified the same problem and also looked at the literature to examine the socio-economic effects from electricity access. Bos, Chaplin, and Mamun (2018), for example, assess the benefits of expanding grid electricity in African countries and Jimenez (2017) studies the development effects of rural electrification more broadly. While both papers find positive welfare effects from electricity access, they also highlight considerable variation in these effects across the surveyed literature. Answering calls for further research that can help us understand the sources of this heterogeneity in impacts from electricity access (Jimenez, 2017, 14), our paper makes two important contributions. First, instead of relying on an informal literature review, we use a systematic review of the literature—an algorithmic procedure to identify existing studies that meet a pre-specified set of inclusion criteria—to provide us with a complete set of papers on household-level impacts from electricity access. Second, and in keeping with Bos, Chaplin, and Mamun (2018, 69) and Jimenez (2017, 6), who emphasize the importance of rigorous evaluation methods, we argue that the empirical strategy, and specifically whether a study uses random assignment of electricity access or not is critical.

In this study, we report results from a systematic review of available studies on the impacts of household electrification in the developing world. We find that as few as 31 impact evaluation studies (out of 7,247 "energy access" studies published since 2000) have been conducted using formal statistical tests, and an even smaller subset of only 7 studies rely on experimental methods, such as for example randomized controlled trials. The analysis shows that studies with experimental methods report fewer positive impacts than observational studies, which we take to include correlational regression, instrumental variable, and difference-in-differences designs. These observational studies, in which assignment to treatment is not random, find positive impacts in 84% of the 44 assessed outcomes, whereas experimental studies only do so 52% of the time for 21 assessed outcomes. In other words, observational studies are roughly 1.6 times more likely to find a positive impact than experimental studies. While we do find differences within the group of observational studies, that is, among descriptive regression, instrumental variable, and difference-in-differences methods, the difference between observational and experimental methods remains the most pronounced.

Notwithstanding some data limitations, we do examine alternative explanations and find that the association between method type and how positively impacts are assessed holds for grid and offgrid studies, different outcome types, and for African and non-African countries. Although we cannot say that method type *causes* fewer positive impacts, our results do highlight the need to improve impact evaluation in electricity access research because policymakers need more rigorous social science research to identify when and how household electrification can generate positive impacts that justify the investment.

2. Impact evaluation and electrification access

Impact evaluation, defined as "an effort to understand whether the changes in well-being are indeed due to project or program intervention," is fundamental to successful policy implementation (Khandker, Koolwal, and Samad, 2009, 18). Quality impact evaluation informs policymakers on which interventions work and how to implement them. By illuminating factors that result in obstacles to success, impact evaluation also contributes to global knowledge that can be used to design effective policies for application to other populations or countries (Gertler et al., 2010).

Identifying cause and effect relationships between a policy and its outcomes is not a trivial problem. As illustrated by the discourse on the "credibility revolution" in development economics, the main impediment to assessing causal effects is finding appropriate counterfactuals (Angrist and Pischke, 2010). Assessing causal effects requires comparing the outcome of interest in those receiving intervention to the outcome that would have happened in the absence of intervention. Yet, a 'true' counterfactual is always unobserved, since no individual or group can exist simultaneously in a state of having experienced and not experienced the policy intervention. Thus, the main challenge to establishing causality is how to develop a comparison group that best mimics the counterfactual.

To address this problem, various methods have been introduced. Broadly, we could categorize them into two groups—observational and experimental (randomized controlled trials). To estimate causal effects using observational data, researchers have employed a range of research designs including difference-in-differences, matching and instrumental variables (Bernard, 2010; Ravallion, 2001). While these estimates can be powerful tools to control for both observable and unobservable confounders, they require detailed data, which is not always readily available, and rely on assumptions to minimize concerns about whether the identified relationship is really causal (Gertler et al., 2010).

On the other hand, randomized controlled trials overcome the inference problem through random assignment. In the randomized experiment, subjects are randomly assigned into treatment and control groups. Random assignment ensures that the only difference in attributes between the two groups are due to random chance, except only for the treatment (intervention). Thus, the difference in outcomes after the intervention should be attributed to the treatment itself, which represents the *causal* effects of treatment.

Although impact evaluations based on randomized controlled trials have been increasingly common in evaluating development outcomes such as public health and education (Bernard, 2010), research on the impacts of electricity access is in its infancy, despite the high stakes and the trillions of dollars being budgeted for the next two decades. Sustainable energy access is widely believed as a main requisite for economic and social development. By allowing the energy-poor to benefit from lighting, water pumps, and modern telecommunication technology, electricity access can directly lead to improvement in various dimensions of development and enhance the quality of life (Lambert et al., 2014; Malakar, 2018). For example, lighting can improve the study environment for children, extend hours for work, and create security. Massive investment in household electricity access, however, comes with opportunity costs. Some scholars have suggested that, for some countries, this money might be better spent on other infrastructure and education programs with larger impacts on economic growth and educational attainment (Lee, Miguel, and Wolfram, 2018; Burlig and Preonas, 2016). There are also concerns that, absent large and continuous government subsidies, electrification of poor households will strain already struggling public utilities, and arguments about whether grid extension or more distributed technologies (mini-grids, solar lanterns) are a better investment.

This is not to say that randomized controlled trails are without problems (Deaton and Cartwright, 2018). They have recently come under fire, in particular over concerns of lack of external validity. External validity is essential when randomized controlled trails are to inform policy (Peters, Langbein, and Roberts, 2018). Studying the trade-off between internal and external validity, Pritchett and Sandefur (2015) show that observational studies within context outperform experimental studies from another context. This is particularly important for our understanding of whether and how results from impact evaluations travel across contexts and countries (Peters, Langbein, and Roberts, 2018; Hamburger et al., 2019).

Whatever the exact method, thin knowledge on the impacts of electricity access policies impedes the formulation of effective interventions and implementation. Without a robust understanding about the costs and benefits of electricity access, policymakers can hardly be informed which policy implementation is the most effective given limited resources. While evidence-based knowledge is not a sufficient condition to develop effective policies, the lack of a solid base of empirical knowledge makes policy failure more likely.

3. Data and methods

Our systematic review summarizes and evaluates the universe of impact assessments focused on *household* electrification in the developing world. Systematic reviews, a cornerstone of medical research (Cochrane Collaboration, 2008), are a process for collecting and analyzing studies on the same (or similar) topics. Systematic reviews differ from the more common "literature review" in several ways. While literature reviews use informal or subjective methods for collecting and analyzing previous research, a systematic review attempts to select and analyze previous studies using a systematic and, as far as possible, replicable method. Systematic reviews hinge on having a focused research question centered around a particular intervention, a specified eligibility criterion for inclusion/exclusion of studies, a systematic search strategy, and an attempt to systematically record process and results of included studies.

In our case, to conduct this review of electrification impact assessments, we began with a literature search. We then developed a coding scheme for the reported impacts from electrification access in these studies, with a focus on five core outcomes: business creation, education, energy expenditure, household income/expenditure, and household savings. For each study, we also classified the method that was used in the analysis, which allows us to explore how the variation in assessed impacts correlates with the observational and experimental methods.

3.1. Literature search: Systematic review of impact evaluations

The literature search proceeded in both a top-down and bottom-up fashion. For the top-down search, we used the following search string in Google Scholar, ScienceDirect, and Web of Science for articles written since 2000: ("rural electrification" or "electricity" or "off-grid" or "solar") AND ("impact" or "effect" or "development" or "benefits"). We focused our search on these main sources as they generally encompass more specific ones, such as JSTOR, EBSCO, or ProQuest. This produced 7,247 articles. To avoid publication bias, we included published, gray and unpublished articles for consideration. From these articles, a large number were excluded based on title and abstract—these were studies that dealt with financing off-grid systems, policies to enable higher adoption of off-grid electrification, and other studies that were not focused particularly on the *impacts* of electricity in developing countries.

As a next step, we evaluated the remaining 514 articles to only include those that conducted some type of evaluation of the effects of electrification on development. We also excluded those studies that were primarily qualitative or reviewed other studies. This left us with 32 articles. Finally, for this study, we narrowed the studies to only those that conducted explicit statistical hypothesis testing of the effects of electrification on one of the following outcomes: energy expenditure, total income/expenditure, savings, business creation, and education—all measured at the household level. This left us with 15 articles from the top-down search.

We offer a detailed description of our filters and show by how much each filter reduces sample size at each stage in Appendix SI1. To reassure the comprehensiveness of our search protocol, we illustrate how our filters work and how they help exclude studies from our final sample. Bastakoti (2006), for example, studies socio-economic impacts from rural electrification in Nepal, but does not apply statistical hypothesis testing. Tasciotti (2017), on the other hand, uses econometric tests to assess impacts from electricity access in Malawi, but is interested in malaria rates as main outcome variable and not socio-economic impacts. Kanagawa and Nakata (2008) and Niu et al. (2013) study educational and socio-economic outcomes of electrification access, as we do, but at the state- and country-level rather than the household level.

The bottom-up search relied on reviews and bibliographies discovered in the top-down search, bibliographies from published articles and books, the authors' own knowledge of the field, and discussions with other authors in the field. From these articles, we discovered 15 articles that passed the topic and methodology restrictions for inclusion. As our unit-of-analysis is an impact evaluation and not a publication, the work by Cook et al. (2005), an Asian Development Bank report, yields two units, one study on Thailand and one on India. Overall, our analyzed sample comprises 31 observations: 24 of these (77%) are peer-reviewed and 7 (23%) are unpublished work, i.e., either working papers or policy reports by, e.g., the World Bank (Khandker et al., 2014) or the Asian Development Bank (Cook et al., 2005). A full list of studies is in Appendix SI2.

Our overall set of studies which we analyze includes 31 studies. Importantly, because we use a systematic review to identify these studies, the small number of studies—an interesting result itself as we discuss in greater detail below—constitutes the complete population of all studies *conditional* on the filters we chose. We do therefore not run into problems of sampling error despite a small sample as our systematic review does not sample but filter. Obviously, using different filters in the systematic review results in a different set of studies to analyze. Here, our choice of filters was motivated by finding research that can both demonstrate that a policy intervention works and can also quantify the size of the effect. In the context of demands for universal energy access, for example in SDG 7, we argue that our filters help identify robust impact evaluations with the potential to better understand energy poverty around the world.

3.2. Coding of assessed impacts

Research on the impacts from rural electrification has assessed a vast array of outcomes. For our purposes, we focus on five broad groups of outcomes at the *household level* that have regularly been assessed: energy expenditure, total income/expenditure, savings, business creation, and education. In selecting these impacts we have been guided by two considerations. First, we chose outcomes that have been analyzed commonly enough so that we can analyze a range of studies. Second, studies that assess these five outcomes have varied in the type of methods that were used to assess impacts. However, this choice also has its limitations. We only consider impacts for households and therefore ignore impacts from electrification that may accrue at the firm, village or regional level (e.g., Peters, Vance, and Harsdorff, 2011; Rud, 2012; Lipscomb, Mobarak, and Barham, 2013), which protects us from introducing 'level of analysis' as a confounding variable. Obviously, we cannot assess all outcomes that there are, so our findings are silent, for instance, about the health effects from rural electrification.

For all the five outcomes we assess, we code their impact as either 'positive' (48 cases), 'neutral' (16 cases) or 'negative' (1 case). For this, we evaluate the impact based on the *main specification* in each paper and follow the studies' authors interpretation of their results. In Appendix SI3, we list how we coded all impacts, together with point estimates and whether they are statistically significant. Positive/negative impacts are those where energy access improves/worsens a household's welfare from a given outcome. If study hours go up, this is a positive impact, whereas a decrease in labor market participation would be coded as a negative impact. Neutral impacts are statistically indistinguishable from zero. For neutral impacts we verified that, for most studies, failure to become statistically significant results from point estimates close to zero rather than imprecisely estimated standard errors. For positive impacts, we base our coding mainly on whether a result is statistically significant or not and hence avoid judgments of whether the impact is large or small. Since cultural, country, and policy contexts as well as how impacts are measured vary dramatically across the studies in our data, making meaningful comparisons of effect sizes is far from trivial, as others have acknowledged (Bos, Chaplin, and Mamun, 2018; Jimenez, 2017; Peters and Sievert, 2016). These comparisons are difficult as only few studies report detailed enough descriptive statistics that would allow us to standardize effect sizes. Our coding is too coarse to capture effect sizes, but—given the limitations in the raw data—relies on establishing if any impacts exist or not, independent of their size. We do also not factor in the cost of providing energy access and stay away from assessing whether an impact, especially if it was fairly expensive to produce, can be considered a success. This is not to say that these aspects are unimportant for impact evaluations of energy access, as we know they are (e.g., Peters, Sievert, and Toman, 2019), yet these assessments are difficult to make in a transparent, replicable, and comparable way across the number of studies in our data.

Another limitation of our coding scheme in this respect is that some studies use several outcomes to conduct a welfare analysis. Lee, Miguel, and Wolfram (2018), for example, assess household income, savings, and education (which we all code as 'neutral') and infer that rural electrification in Kenya has been welfare decreasing because the consumer surplus from electricity access falls below total construction cost. Importantly, our study does not even get into welfare analysis as our goal is much more modest. We do not assess broader conclusions (which may be prone to authors' subjective judgment on the importance of outcomes) of the studies in our data, but really rather evaluate the impacts outcome by outcome. For us, evaluating impacts before any cost/benefit analysis takes place is important in order to answer the more basic question of whether any impacts from rural electrification exist at all—without any benefits it is hard to justify any cost. Moreover, only five studies in our data explicitly attempt a welfare analysis, so any comparisons on this point would be really tenuous.

We next offer some details on the five outcomes which we assess in the studies identified from our systematic review. In coding impacts, our first category, *energy expenditure* captures households' energy cost burdens which affect the allocation decision of scarce resources to energy spending (Alkon, Harish, and Urpelainen, 2016). We coded an impact as positive if a study finds decreases in fuel expenditures. Also, we include household expenditures on specific fuels such as LPG, electricity and kerosene. Some studies focus specifically on expenditures on kerosene which is the main source of backup lighting where electricity is not reliably accessible (Kudo, Shonchoy, and

Takahashi, 2017; Lee, Miguel, and Wolfram, 2018).

The next three categories, total household income/expenditure, savings and business creation, directly speak to the poverty-alleviating effects of household electrification. There is a common suggestion among policy communities and academic scholars that electricity access policy such as solar home systems and micro-grids lead to increases in the household-level economic welfare, especially in income or expenditure (Bensch, Peters, and Schmidt, 2012; Aklin et al., 2017).

In the *household income* category, we also include studies with indirect measures of the household-level economic benefits, such as studies using employment status (Grogan, 2016), male and female labor market participation (Dinkelman, 2011), and total hours worked (Lee, Miguel, and Wolfram, 2018). We do so as it is very likely that these measures will lead to higher household income over time. In addition to household income, a handful of studies use household-level *savings* as well as total asset values as a measure of economic benefits from electrification (Khandker et al., 2014; Aevarsdottir, Barton, and Bold, 2017). Next, we treat *business creation* as a separate category for how improved lighting, for example, can increase economic activity (Aklin et al., 2017). More generally, access to better lighting and modern fuels can have intermediate effects on business practices and profits through extending working hours for small businesses after dusk (Akpan, Essien, and Isihak, 2013).

Finally, we focus on *education* as one of our core impact outcomes. Electrification is expected to have an independent and complementary effect on educational opportunities (Khandker, Barnes, and Samad, 2013). To measure the educational effects, studies rely on indicators ranging from study hours (Agoramoorthy and Hsu, 2009) and academic achievements (Kudo, Shonchoy, and Takahashi, 2017) to school enrollment rates (Khandker, Barnes, and Samad, 2013). Our classification of *education* encompasses all these indicators.

The discussion above shows that different studies operationalize the same impacts in different ways. To capture this variation, we distinguish between more and less direct operationalizations of the assessed impact. A study that assesses the impact on household income and also uses a direct measure of household income as its outcome variable is a direct operationalization, whereas measuring labor market participation would only be an indirect measure of household income. Table 1 shows that 71% of the assessed outcomes (46 out of 65) are direct measures, with less than a third being measured more indirectly.

Outcome	Coding		
	Total	Direct	Indirect
Energy expenditure	16	9	7
Total income/expenditure	22	15	7
Savings	5	3	2
Business creation	3	1	2
Education	19	18	1
Total	65	46	19

Table 1: Direct and indirect measures of impact outcomes. Columns (2), (3), and (4) show the total number of assessed outcomes as well as counts of the direct and indirect measures for the five assessed impact outcomes.

The largest variation in how impacts from rural electrification are measured occur in the household income and energy expenditure categories. This is not too surprising as not only are these among the most frequently assessed outcomes, but, as mentioned before, employment status and labor market participation do often proxy for household income. Similarly, energy expenditures are often measured for different fuels, such as LPG or kerosene, rather than total expenditures, which drives variation.

3.3. Coding of study method

Impact assessments not only vary in the type of impacts they assess, but also in the type of methods they use to produce these results. We developed a coding scheme to classify these methods. In a broad distinction, we first group studies into whether they use "observational" or "experimental" methods. We classified a study as experimental if and only if it used random assignment to treatment and control, as it is done for instance in randomized controlled trials where subjects (e.g., households, villages, regions) are randomly assigned to experimental conditions. Otherwise, we classified a study as observational.

For a more fine-grained analysis, we further break out the observational category into "descriptive" studies, "instrumental variable" (IV) studies, and "difference-in-differences" (DID) studies. We classify a study as descriptive if it uses, for example, mean comparisons or regression models to statistically test a hypothetical relationship between electrification and an impact, without much concern about causal identification. To identify IV studies, we check if, for the main model specification, a two-stage IV estimation is clearly stated. For DID methods, we require a study to use (i) panel data and (ii) include two-way unit and time fixed effects. Table 2 summarizes the four methods types into which we group the studies form our data.

Method		Criterion	N
Broad	Detailed		
Observational	Descriptive	Mean comparisons or regressions	10
	IV	Two-stage least squares (2SLS) estimation	8
	DID	Panel data analysis with unit and time fixed effects	6
Experimental	Experiment	Researcher-controlled randomization	7

Table 2: Coding of four different types of methods in impact assessment studies. Columns (1) and (2) provide information on the different methods we distinguish, column (3) specifies the coding criterion, and column (4) reports the number of studies that use a given method in our data. Sample size $N = 31$.

Many studies report results from several different analyses using different model specifications and research designs. For our coding, we rely on the *main* model specification that is used for interpreting the study's key findings. For example, our data includes a study that mainly uses DID estimation but also employs IV estimation as a robustness check (Rao, 2013). As the main models use difference-in-differences, we code it as such, and not as an IV study. Similarly, there are five studies that mainly use IV estimation as main model specification, while also estimating fixed effects models (Chakravorty, Pelli, and Marchand, 2014; Dinkelman, 2011; Khandker, Barnes, and Samad, 2012; Grogan, 2016; van de Walle et al., 2017). We classify these studies as IV studies, following the same logic.

Before proceeding to present our empirical findings, let us highlight, as a word of caution, that some may find our main distinction into experimental and observational studies too crude and would prefer a more nuanced distinction of methodological differences along the lines of how rigorously a study was conducted or how properly it was implemented. While we are sympathetic towards such a distinction at a conceptual level, measuring a study's rigor in an objective and replicable fashion is almost impossible. Instead, we side with our distinction into experimental and observational studies, which we think is appealing in two important ways: first, it is fairly straightforward to operationalize; second, we think it captures the notion of a study's rigor sufficiently well, albeit imperfectly.

4. Results

Our analysis produces two clear results, based on the descriptive patterns in the data. First, very few impact evaluation studies of electricity access exist. Since 2000, we could only identify 31 studies from our systematic review that meet our inclusion criterion of using statistical hypothesis testing—or, 1.6 studies per year. With 19 studies having been published within the last five years, the situation is somewhat improving, but the absolute number remains tiny. In order to offer evidence-based policy recommendations, more studies are needed. In order to assess and compare effect sizes of policy interventions of electricity access, outcome measures need to be better standardized and basic descriptive statistics need to be reported to allow for a rigorous meta-analysis (Jimenez, 2017). For instance, in our data, 19 studies evaluate educational impacts, yet only 10 studies use some version of a study time measure (although on different scales), while others measure educational impacts as completed years of schooling, literacy rates, or test scores. We offer a detailed breakdown of the analyzed studies by country, technology, and what outcomes are covered in Appendix SI3.

Second, and more worrying for energy policy, among these few studies we find that those using experimental methods are associated with fewer positive impacts, even though there is also important variation among observational studies. Overall, this is first evidence to caution against overly optimistic expectations of impacts from rural electrification for socio-economic development. Our findings also highlight the importance of recognizing the method that was used in an impact assessment when interpreting study results and making inferences about the generalizability of the results.

We report our results in three steps. We first only distinguish between observational and experimental methods. In a follow-up analysis then, we break out observational methods into descriptive, IV, and DID studies to highlight variation within the group of observational methods. Finally, we consider some alternative explanations for the patterns in the data.

4.1. Results for observational versus experimental methods

The main insight from comparing observational and experimental methods is that the type of method matters. Studies based on experimental methods find fewer positive impacts than observational studies. The mosaic plot in Figure 1 illustrates this.

Overall, the 31 studies in our sample assess 65 impacts. The vast majority of impacts, i.e., 48 are positive, while 16 studies find neutral results; only one study finds a negative impact from electricity access, namely a reduction in female labor market participation because of electricity access (Grimm, Sparrow, and Tasciotti, 2015). With 24 observational studies and only 7 experimental ones, the share of positive impacts in observational studies (84%) is much higher than in experimental work (52%).

One possibility to explain this finding is that different methods focus on different technologies, and grid access certainly produces more positive impacts than offgrid electricity access. Most of our understanding of impacts from electricity access relies on studies that focus on grid access: 22 studies (71%) in our sample study impacts from grid electrification and 9 studies (29%) study impacts from offgrid technology. Indeed, grid/offgrid technology is strongly correlated with method

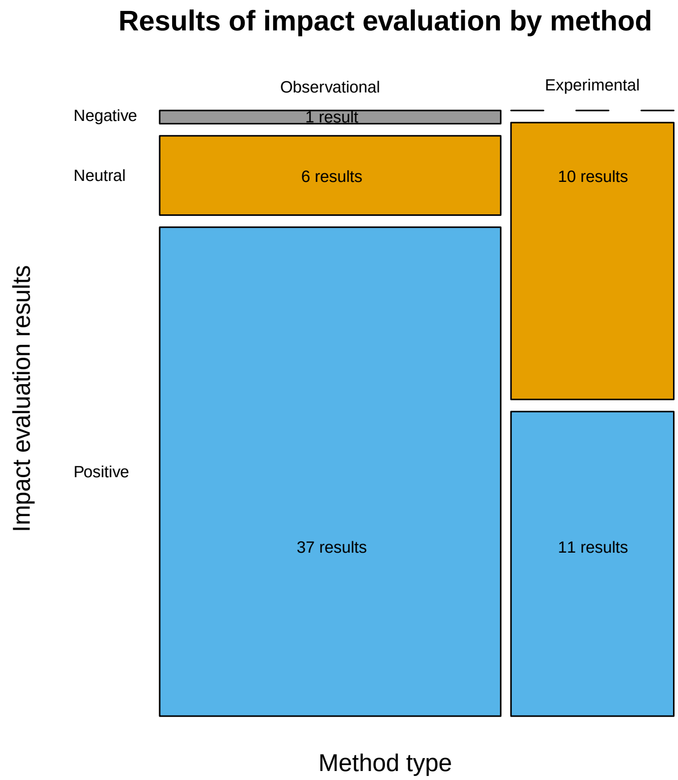

Figure 1: Mosaic plot. The plot shows the number of positive (blue), neutral (orange), and negative (gray) impacts in the assessed studies ($N = 31$) by observational and experimental method. The dashed line indicates that no experimental study found negative impacts.

type because 90% of studies focusing on grid access use observational methods, whereas only 44% of offgrid studies do so. This poses an obvious empirical challenge for our analysis: fewer positive impacts in experimental studies could be a result of the chosen method, but could also come from the fact that these studies predominantly assess offgrid systems, which promise smaller impacts to begin with.

However, Figure 2 breaks the mosaic plot from above out by grid and offgrid studies and a similar descriptive pattern prevails. Holding technology constant, among grid studies (left panel) the share of positive impacts from observational studies is 85% compared to 40% in experimental work. Similarly, for studies assessing offgrid electricity access (right panel), positive impacts occur 80% of the time in observational studies, but only 56% of the time in experimental studies. Even in the subsamples for identical technologies, the ratios and descriptive patterns are roughly the same as in the complete data.

Fully separating technology effects from method effects is nonetheless challenging because of the above described separation issue in our fairly small data set. In the supplementary information SI5 we show results from regression models, which we think are a useful, albeit imperfect, way to "net out" the grid/offgrid technology effect. Point estimates suggest that experimental studies

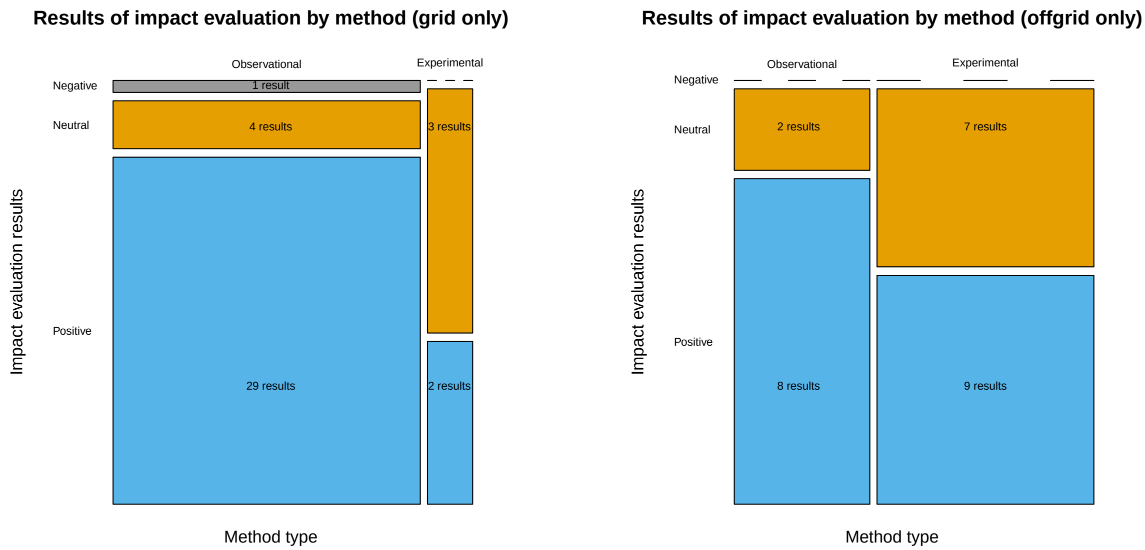

Figure 2: Mosaic plots, separately for grid (left panel) and offgrid studies (right panel). Each plot shows the number of positive (blue), neutral (orange), and negative (gray) impacts in the assessed studies ($N = 31$) by observational and experimental method. The dashed line indicates that no experimental study found negative impacts.

are more likely to find positive impacts than non-experimental studies but results are mostly not statistically significant at conventional levels. Considering limitations of our data, the findings from our different types of descriptive analyses provide preliminary evidence that empirical strategy is indeed correlated with how positively impacts from electricity access are evaluated. Since none of these findings should be understood as causal relationships, we discuss alternative explanations that could produce a similar pattern in the data below.

Figure 3 shows how impacts can be broken down across the five outcomes we study. This is insightful as it demonstrates that the difference in positive and neutral impacts across outcomes is largest for household income and educational outcome variables. Clearly, these are the outcomes that are assessed the most, accounting for almost two-thirds of our data, but the differences across observational and experimental methods are stark. The shares of positive findings for household income impacts are 82% for observational and 25% for experimental studies; descriptively, evidence is even stronger for educational impacts where all assessed impacts are positive for observational studies, while only 33% of experimental studies find positive impacts.

4.2. Results for all four method types

Our analysis, so far, presented results for separating method type into observational and experimental methods. We now differentiate further and break out the observational category into descriptive, IV, and DID studies, as described in our methods section above. In Table 3 we show how negative, neutral, and positive impacts are distributed across these four, more nuanced methods types. To summarize this data in a useful way, we calculate the share of positive impacts over total assessed impacts in percent, shown in the last column.

With an average share of 73% across all studies, we find that descriptive (88%) and IV methods (90%) produce consistently above-average positive impacts, while DID (75%) and experimental

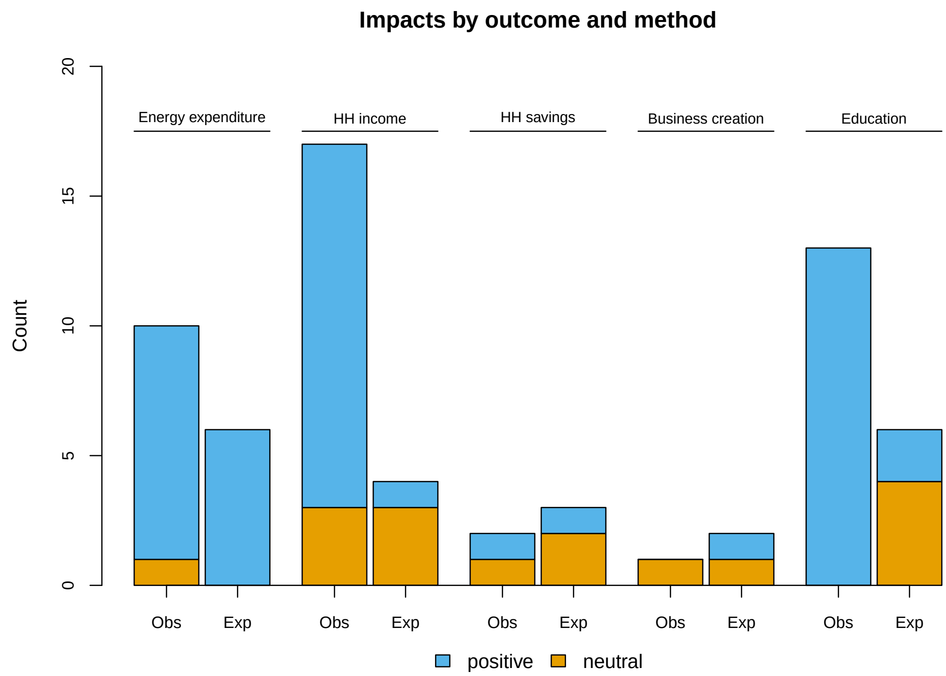

Figure 3: Bar plot. The bar plot shows positive and neutral results by outcome variable. For presentation, we exclude one observational study that produces a negative impact. The shares of positive impacts for observational and experimental methods are: Energy expenditure: 90% and 100%; Income: 82% and 25%; Savings: 50% and 33%; Business creation: 0% and 50%; and Education: 100% and 33%.

Study characteristic		Result of impact evaluation			Impacts #	Share of positive impact
Method	Detailed method	Negative	Neutral	Positive		
Observational	Descriptive	0	2	15	17	88%
	IV	0	1	10	11	90%
	DID	1	3	12	16	75%
Experimental	Experiment	0	10	11	21	52%
Total		1	16	48	65	73%

Table 3: Results of impact evaluations in assessed studies ($N = 31$) by method type. Columns (3)-(5) show counts of negative, neutral, and positive results of the impact evaluations. Column (7) calculates the share of positive impacts over the total assessed impacts for each method in percent.

methods (52%) have lower shares. This variation within the observational methods is interesting for two reasons. First, IV studies do stand out as such studies that almost always produce positive impacts; they also assess a smaller number of impacts than other studies, for an average of 1.4 impacts per study—compared to 1.7 for descriptive studies, 2.6 for DID, and 3 for experimental studies. Either way, IV studies might follow a somewhat different logic in impact assessment of electricity access research. Second, DID studies seem to be different from the other types among the set of observational studies. They are in several regards much more similar to experimental studies than to descriptive or IV methods. Given a greater emphasis on causal identification in

DID designs, this similarity may be not too surprising and corresponds well to the hierarchy in evaluation methods (Jimenez, 2017).

In the appendix SI5 we again report results from several regression models. As before, there is some evidence that experimental methods are mostly correlated with fewer positive impacts than other methods although only rarely at conventional levels of statistical significance. We take these results to be highly suggestive of a descriptive pattern in the data: experimental methods produce fewer positive impacts compared to other methods. Given how difficult it is to address the selection bias in electricity access research (e.g., Bos, Chaplin, and Mamun, 2018, for an excellent discussion), identifying the *true* effect is challenging, absent any random assignment. Typically, households that get access to electricity first live in more developed and easier-to-electrify communities, which produces exactly the positive/upward bias in impact evaluations that cannot perfectly purge all the confounding effects.

4.3. Alternative explanations

The main point of our paper has been to show that the type of method a study uses to assess impacts from electricity access is correlated with how positively it evaluates these impacts. Obviously, since our analysis is correlational, we cannot say that the differences in impacts is *caused* by method type, so probing alternative explanations is important. Above, we pointed to grid/offgrid technology as a potential confounder in this relationship. Another reasonable concern might be that how 'ambitious' the outcome of an impact assessment is may matter for how positively a study evaluates the impacts. If this ambition level is correlated with method type, it may produce the same pattern in the data as we describe it, but for different reasons.

In order to probe this alternative explanation, we separated our outcomes into less and more ambitious ones. Notably, this is not a normative distinction, but rather one along the lines of how easily or immediately welfare gains from electricity access materialize. Getting access to nighttime lighting enables children to study more at night right away, while opening a business as a result of electricity access may take some time. Following this logic, we classify impacts related to education and energy expenditures as more immediate, while changes to household income, household savings, and business creation are more ambitious in that sense.

Holding the type of outcome constant, we do again find strong descriptive evidence that method type matters. For "less ambitious" outcome variables, observational studies find positive impacts 71% (15 out of 21) of the time over only 33% (3 out of 9) in experimental studies. For the more long-term outcomes, such as increases in household income, almost all observational studies find positive impacts (95%, 22 out of 23) relative to only 67% (8 out of 12) in experimental studies. Given the data we have, these results make us confident that differences in ambition levels are unlikely to produce the pattern in the data that we attribute to method type. This discussion however also calls for a better understanding of not only how persistent impacts are over time but also when one should expect these impacts to materialize.

Another recurring finding in the literature is that impacts from electricity access are generally smaller in Africa than in South East Asia and Latin America (Peters and Sievert, 2016; Bos, Chaplin, and Mamun, 2018; Hamburger et al., 2019). It is hence useful to check how our argument about the importance of method type fares when splitting our data by region. Twelve studies in our data, or a bit more than a third, cover seven African countries, assessing 27 impacts, 18 of which are positive and 9 are neutral. As before, we do find the same patterns in the data for African and non-African countries. Outside Africa, observational studies assess 26 out of 30 impacts positively (87%), but experimental studies find only 4 out of 8 positive impacts. In Africa, consistent with the literature, the relative share of positive outcomes is indeed much smaller (67%) compared to non-African states (79%), but this does not break the link between method type and positive impacts. Observational studies assess impacts positively 78% of the time, but experimental studies do so only 54% of the time. Our main finding therefore holds similarly and at comparable rates for African and non-African regions.

5. Conclusion and policy implications

Despite data limitations, which clearly warrant additional research about sources of heterogeneity in impacts from rigorous impact assessments of electricity access, the main message of our study is clear. There are only few studies available that assess impacts from electricity access, and even fewer that do so with experimental methods. This paucity of experimental studies is possibly problematic. We show descriptive evidence that demonstrates that experimental studies are correlated with consistently fewer positive impacts. While it is true that observational studies prioritize assessing impacts from grid access whereas experimental studies assess mostly offgrid technologies, which obviously complicates the analysis, our findings should not be readily dismissed. Slicing and dicing the data in multiple ways, the same pattern of experimental studies evaluating impacts less favorably emerges over and over again—for grid and offgrid studies, for different types of outcome variables, and inside and outside Africa.

Albeit preliminary because of the very nature of our data, we think these findings are an important wake-up call for both electricity access research and practice. In interpreting findings from impact evaluations, we should carefully consider what method had been used in these studies. As others have noted (Bos, Chaplin, and Mamun, 2018; Jimenez, 2017), assessing impacts from electricity access is plagued by selection problems, and these do not work in favor of helping policymakers to rigorously assess whether an electricity access policy works or not. Basing their decisions and the design of policy interventions on results from observational studies that are probably less ideal for impact evaluation, the evidence base for policy remains uncertain.

Our research shows that impact evaluations from observational studies tend to over-promise positive impacts, as compared to randomized controlled trials that rely on randomization to produce estimates of the causal effect of any given policy intervention. Experimental studies, despite their strengths that derive from causal identification, face other limitations. Some of the biggest concerns have been the lack of external validity (Peters, Langbein, and Roberts, 2018) and the worry that an unjustified preoccupation with experimental methods distorts research away from the most important questions to those which can be studied using an experimental paradigm (Ravallion, 2018).

It is true that only two studies in our data set use experimental methods to assess impacts from grid access, but while some might consider this a limitation, it also showcases that impacts from grid expansion can be assessed with randomized controlled trials, too. For this to work, however, more collaboration across policymakers and researchers is needed: on the one hand, to make inferences from observational studies more credible, on the other hand to scale up randomized controlled trials to allow for impact assessment of large-scale grid extension programs. While offgrid technologies may initially be (believed to be) easier to assess with research designs using randomization, coordination between researchers and policymakers can remove this perceived trade-off between a robust methodology and the scale of an intervention. Impact evaluations of large-scale grid extension programs can leverage randomization by, for instance, randomizing the sequence in which households or neighborhoods are electrified. Clearly, the energy-poor of the world deserve policies that have been evaluated for their impacts.

Finally, impact evaluation research should focus on identifying the conditions under which interventions generate substantial benefits. The development literature has long been emphasizing the role of the local context (Ostrom et al., 1999; Aklin et al., 2018), so identifying and understanding how these local factors condition the impacts of electricity access interventions will prove critical for eradicating energy poverty. Our findings call for an honest reassessment of what we really know, in research and practice, about the impacts of electricity access.

Data availability

The data set used to produce the results of this paper and all associated code will be made available from PB's Harvard Dataverse at `https://dataverse.harvard.edu/dataverse/patrickbayer` upon publication.

Acknowledgements

We thank Todd Moss, Jörg Peters, and S.P. Harish for comments on an earlier version of the paper. Rebecca Yim proofread the manuscript.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Aevarsdottir, Anna Margret, Nicholas Barton, and Tessa Bold. 2017. "The IImpact of Rural Electrification on Labour Supply, Income, and Health: Experimental Evidence with Solar Lamps in Tanzania." IGC Working paper, E-89302-TZA-1.

Agoramoorthy, Govindasamy, and Minna J Hsu. 2009. "Lighting the Lives of the Impoverished in India's Rural and Tribal Drylands." *Human Ecology* 37 (4): 513–517.

Aklin, Michaël, Patrick Bayer, S.P. Harish, and Johannes Urpelainen. 2017. "Does Basic Energy Access Generate Socioeconomic Benefits? A Field Experiment with Off-grid Solar Power in India." *Science Advances* 3 (5): e1602153.

Aklin, Michaël, Patrick Bayer, S.P. Harish, and Johannes Urpelainen. 2018. *Escaping the Energy Poverty Trap: When and How Governments Power the Lives of the Poor.* Cambridge, MA: MIT Press.

Akpan, Uduak, Maurice Essien, and Salisu Isihak. 2013. "The Impact of Rural Electrification on Rural Micro-enterprises in Niger Delta, Nigeria." *Energy for Sustainable Development* 17 (5): 504–509.

Alkon, Meir, S.P. Harish, and Johannes Urpelainen. 2016. "Household Energy Access and Expenditure in Developing Countries: Evidence from India, 1987–2010." *Energy for Sustainable Development* 35: 25–34.

Angrist, Joshua D, and Jörn-Steffen Pischke. 2010. "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con Out of Econometrics." *Journal of Economic Perspectives* 24 (2): 3–30.

Bastakoti, Badri Prasad. 2006. "The Electricity-Livelihood Nexus: Some Highlights from the Andhikhola Hydroelectric and Rural Electrification Centre (AHREC)." *Energy for Sustainable Development* 10 (3): 26–35.

Bensch, Gunther, Jörg Peters, and Christoph M Schmidt. 2012. "Impact Evaluation of Productive Use–An Implementation Guideline for Electrification Projects." *Energy Policy* 40: 186–195.

Bernard, Tanguy. 2010. "Impact Analysis of Rural Electrification Projects in Sub-Saharan Africa." *The World Bank Research Observer* 27 (1): 33–51.

Bos, Kristine, Duncan Chaplin, and Arif Mamun. 2018. "Benefits and Challenges of Expanding Grid Electricity in Africa: A Review of Rigorous Evidence on Household Impacts in Developing Countries." *Energy for Sustainable Development* 44: 64–77.

Burlig, Fiona, and Louis Preonas. 2016. "Out of the Darkness and Into the Light? Development Effects of Rural Electrification in India." Energy Institute at Haas, Working Paper 268.

Chakravorty, Ujjayant, Martino Pelli, and Beyza Ural Marchand. 2014. "Does the Quality of Electricity Matter? Evidence from Rural India." *Journal of Economic Behavior & Organization* 107: 228–247.

Cochrane Collaboration. 2008. *Cochrane Handbook for Systematic Reviews of Interventions.* London: Cochrane Collaboration.

Cook, Cynthia C, Tyrrell Duncan, Somchai Jitsuchon, Anil Sharma, and Wu Guobao. 2005. *Assessing the Impact of Transport and Energy Infrastructure on Poverty Reduction.* Metro Manila, the Philippines: Asian Development Bank: Asian Development Bank.

Deaton, Angus, and Nancy Cartwright. 2018. "Understanding and Misunderstanding Randomized Controlled Trials." *Social Science & Medicine* 210: 2–21.

Dinkelman, Taryn. 2011. "The Effects of Rural Electrification on Employment: New Evidence from South Africa." *American Economic Review* 101 (7): 3078–3108.

Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel M. J. Vermeersch. 2010. *Impact Evaluation in Practice.* Washington, DC: World Bank: World Bank Publications.

Grimm, Michael, Robert Sparrow, and Luca Tasciotti. 2015. "Does Electrification Spur the Fertility Transition? Evidence From Indonesia." *Demography* 52: 1773–1796.

Grogan, Louise. 2016. "Household Electrification, Fertility, and Employment: Evidence from Hydroelectric Dam Construction in Colombia." *Journal of Human Capital* 10 (1): 109–158.

Hamburger, David, Joel Jaeger, Patrick Bayer, Ryan Kennedy, Joonseok Yang, and Johannes Urpelainen. 2019. "Shades of Darkness or Light? A Systematic Review of Geographic Bias in Impact Evaluations of Electricity Access." *Energy Research & Social Science* 58: 101236.

IEA. 2017. "Energy Access Outlook 2017: From Poverty to Prosperity." World Energy Outlook Special Report.

Jimenez, Raul. 2017. "Development Effects of Rural Electrification." Inter-American Development Bank Policy Brief IDB-PB-261.

Kanagawa, Makoto, and Toshihiko Nakata. 2008. "Assessment of Access to Electricity and the Socio-economic Impacts in Rural Areas of Developing Countries." *Energy Policy* 36 (6): 2016–2029.

Khandker, Shahidur R, Douglas F Barnes, and Hussain A Samad. 2012. "The Welfare Impacts of Rural Electrification in Bangladesh." *The Energy Journal* pp. 187–206.

Khandker, Shahidur R, Douglas F Barnes, and Hussain A Samad. 2013. "Welfare Impacts of Rural Electrification: A Panel Data Analysis from Vietnam." *Economic Development and Cultural Change* 61 (3): 659–692.

Khandker, Shahidur R, Gayatri B Koolwal, and Hussain A Samad. 2009. *Handbook on Impact Evaluation: Quantitative Methods and Practices.* Washington, DC: World Bank: World Bank Publications.

Khandker, Shahidur R., Hussain A. Samad, Sadeque Zubair K. M., Mohammed Asaduzzaman, Mohammad Yunus, and A. K. Enamul Haque. 2014. *Surge in Solar-Powered Homes : Experience in Off-grid Rural Bangladesh.* Washington, DC : World Bank Group.: World Bank.

Kudo, Yuya, Abu S Shonchoy, and Kazushi Takahashi. 2017. "Can Solar Lanterns Improve Youth Academic Performance? Experimental Evidence from Bangladesh." *The World Bank Economic Review* .

Lambert, Jessica G., Charles A.S.Hall, Stephen Balogh, Ajay Gupta, and Michelle Arnold. 2014. "Energy, EROI and Quality of Life." *Energy Policy* 64: 153–167.

Lee, Kenneth, Edward Miguel, and Catherine Wolfram. 2018. "Experimental Evidence on the Economics of Rural Electrification." Working Paper.

Lenz, Luciane, Anicet Munyehirwe, Jörg Peters, and Maximiliane Sievert. 2017. "Does Large-Scale Infrastructure Investment Alleviate Poverty? Impacts of Rwanda's Electricity Access Roll-Out Program." *World Development* 89 (Supplement C): 88 – 110.

Lipscomb, Molly, A. Mushfiq Mobarak, and Tania Barham. 2013. "Development Effects of Electrification: Evidence from the Topographic Placement of Hydropower Plants in Brazil." *American Economic Journal: Applied Economics* 5 (2): 200–231.

Malakar, Yuwan. 2018. "Evaluating the Role of Rural Electrification in Expanding People's Capabilities in India." *Energy Policy* 114: 492–498.

Niu, Shuwen, Yanqin Jia, Wendie Wang, Renfei He, Lili Hu, and YanLiu. 2013. "Electricity Consumption and Human Development Level: A Comparative Analysis based on Panel Data for 50 Countries." *International Journal of Electrical Power & Energy Systems* 53: 338–347.

Ostrom, Elinor, Joanna Burger, Christopher B. Field, Richard B. Norgaard, and David Policansky. 1999. "Revisiting the Commons: Local Lessons, Global Challenges." *Science* 284 (5412): 278–282.

Peters, Jörg, Colin Vance, and Marek Harsdorff. 2011. "Grid Extension in Rural Benin: Micro-manufacturers and the Electrification Trap." *World Development* 39 (5): 773–783.

Peters, Jörg, Jörg Langbein, and Gareth Roberts. 2018. "Generalization in the TropicsâĂŞDevelopment Policy, Randomized Controlled Trials, and External Validity." *World Bank Research Observer* 33 (1): 34–64.

Peters, Jörg, and Maximiliane Sievert. 2016. "Impacts of Rural Electrification Revisited: The African Context." *Journal of Development Effectiveness* 8 (3): 327–345.

Peters, Jörg, Maximiliane Sievert, and Michal A. Toman. 2019. "Rural Electrification through Mini-Grids: Challenges Ahead." *Energy Policy* 132: 27–31.

Pritchett, Lant, and Justin Sandefur. 2015. "Learning from Experiments When Context Matters." *American Economic Review* 105 (5): 471–475.

Rao, Narasimha D. 2013. "Does (Better) Electricity Supply Increase Household Enterprise Income in India?" *Energy policy* 57: 532–541.

Ravallion, Martin. 2001. "The Mystery of the Vanishing Benefits: An Introduction to Impact Evaluation." *The World Bank Economic Review* 15 (1): 115–140.

Ravallion, Martin. 2018. “Should the Randomistas (Continue to) Rule?” Center for Global Development Working Paper No. 492.

Rud, Juan Pablo. 2012. “Electricity Provision and Industrial Development: Evidence from India.” *Journal of Development Economics* 97 (2): 352–367.

SE4ALL. 2017. “Why Wait? Seizing the Energy Access Dividend.” Available online at https://www.seforall.org/sites/default/files/Why_Wait-Full.pdf. Accessed on 25 June 2018.

Tasciotti, Luca. 2017. “Use of Electricity and Malaria Occurrence: Is There a Link? The Case of Malawi.” *Energy Policy* 101: 310–316.

UNDP, and UN Environment. 2018. “Policy Brief #5 on Financing SDG.” Prepared in support of SGG7 review at the UN High-Level Political Forum on Sustainable Development, July 2018. Available online at https://sustainabledevelopment.un.org/content/documents/17549PB_5_Draft.pdf. Accessed on 25 June 2018.

van de Walle, Dominique, Martin Ravallion, Vibhuti Mendiratta, and Gayatri Koolwal. 2017. “Long-term Gains from Electrification in Rural India.” *World Bank Economic Review* 31 (2): 385–411.