Ankel-Peters, Jörg; Schmidt, Christoph M.

Working Paper

Rural Electrification, the Credibility Revolution, and the Limits of Evidence-Based Policy

I4R Discussion Paper Series, No. 123

Provided in Cooperation with:
The Institute for Replication (I4R)

Suggested Citation: Ankel-Peters, Jörg; Schmidt, Christoph M. (2024) : Rural Electrification, the Credibility Revolution, and the Limits of Evidence-Based Policy, I4R Discussion Paper Series, No. 123, Institute for Replication (I4R), s.l.

This Version is available at:
https://hdl.handle.net/10419/295221

INSTITUTE for REPLICATION

No. 123
I4R DISCUSSION PAPER SERIES

Rural Electrification, the Credibility Revolution, and the Limits of Evidence-Based Policy

Jörg Ankel-Peters

Christoph M. Schmidt

May 2024

INSTITUTE for REPLICATION

I4R DISCUSSION PAPER SERIES

I4R DP No. 123

Rural Electrification, the Credibility Revolution, and the Limits of Evidence-Based Policy

Jörg Ankel-Peters[1,2], Christoph M. Schmidt[1,3]

[1]RWI – Leibniz Institute for Economic Research, Essen/Germany
[2]University of Passau/Germany
[3]Ruhr University Bochum/Germany

MAY 2024

E-Mail: joerg.peters@rwi-essen.de
RWI – Leibniz Institute for Economic Research
Hohenzollernstraße 1-3
45128 Essen/Germany
www.i4replication.org

ISSN: 2752-1931

Rural electrification, the credibility revolution, and the limits of evidence-based policy

Jörg Ankel-Peters[1,2] and Christoph M. Schmidt[1,3]

[1]*RWI – Leibniz Institute for Economic Research,* [2]*University of Passau,* [3]*Ruhr University Bochum*

May 2024

Abstract

The so-called credibility revolution dominates empirical economics, with its promise of causal identification to improve scientific knowledge and ultimately policy. By examining the case of rural electrification in the Global South, this opinion paper exposes the limits of this evidence-based policy paradigm. The electrification literature boasts many studies using the credibility revolution toolkit, but at the same time several systematic reviews demonstrate that the evidence is divided between very positive and muted effects. This bifurcation presents a challenge to the science-policy interface, where policymakers, lacking the resources to sift through the evidence, may be drawn to the results that serve their (agency's) interests. The interpretation is furthermore complicated by unresolved methodological debates circling around external validity as well as selective reporting and publication decisions. These features, we argue, are not particular to the electrification literature but inherent to the credibility revolution toolkit.

Keywords: Energy access, evidence-based decision-making, systematic reviews, meta-science

JEL-Codes: O13, D78

Acknowledgments: We are grateful for valuable comments and suggestions by E. Somanathan, two anonymous referees, Gunther Bensch, Maximiliane Sievert, Colin Vance and to participants of the Sustainable Energy Transition Initiative (SETI) 2020 workshop, International Association of Energy Economics (IAEE) 2021 conference, the Power to Empower Emerging Africa 2020 workshop in Marrakesh and the 3rd Conference on Econometrics and the Environment 2020. All correspondence to joerg.peters@rwi-essen.de

1. Introduction

The extent to which the high costs of rural electrification are justified by its impacts on societies and economies has been a matter of debate for decades (see, for example, Rose 1940, Barnes 2010, Barnes and Binswanger 1986, Devine 1983). In recent years, academic contributions to this discussion have been influenced considerably by the so-called *credibility revolution* in economics (see Angrist and Pischke 2010). The claim is that "design-based research" (Card 2022) like randomized controlled trials (RCTs) and instrumental variables (IVs) leads to more credible and verifiable identification of causal effects. This "experimentalist paradigm" (Biddle and Hamermesh 2017) is closely linked to the vision of evidence-based policy: well-identified causal effects, so the narrative goes, will eventually tell us which interventions work and hence should be scaled to shape future policies (Duflo 2004; Duflo 2020; Panhans and Singleton 2017; Young et al. 2002).

In this paper, we examine the case of rural electrification in the Global South, documenting that design-based research is much less effective in improving policy than it is often claimed. This is not a new verdict, and we build on previous critical reflection on the credibility revolution paradigm (Basu 2014; Deaton 2020; Deaton and Cartwright 2018; Drèze 2020; Heckman and Urzúa 2010; Muller 2023; Rodrik 2008; Ravallion 2009, 2020).[1] We extend this line of discussion by a specific application to rural electrification, an important area of development policy that absorbs large amounts of public funding (Blimpo and Cosgrove-Davies 2019; World Bank 2018). While national governments often justify investments into rural electrification from a social justice and hence a rights-based perspective, donor agencies and international development banks are under pressure to prove that the investment is worthwhile, following an explicit or implicit cost-benefit analysis logic. There is also an interesting within-sector cost-effectiveness debate because investment-intensive extension of the power grid into rural areas competes with infrastructure leapfrogging via lower cost decentralized solutions like stand-alone solar or mini-grids (Levin and Thomas 2016).

To inform this debate, many empirical studies have been published in recent years, increasingly also using design-based methods from the credibility revolution toolkit. Several systematic reviews and meta-analyses have summarized this growing literature. In a nutshell, these reviews show that the literature is divided, with some studies finding very large effects, and others very modest or no effects. This divide is consequential for policy, especially considering that, for the extension of the power grid, large effects are required to justify the high costs. This holds true under a cost-benefit principle as it is applied by many donors, but

[1] A less economics-centric introspection reveals that similar debates about positivist claims for epistemological hegemony have been well-known in the sociology of science for decades (see, for example, Collins and Evans 2002). See also Whittington et al. (2012) for a perspective on the use of evidence in the water and sanitation sector.

also under a rights-based principle because then grid extension competes with off-grid technologies for cost-effectiveness.

Such meta-analyses and systematic reviews are important because, while design-based research is good at generating well-identified causal effects, the external validity gap still needs to be bridged. For this, an accumulation of evidence is needed – something that Duflo (2020) refers to as the "pointillist painting," with each causal study being one dot on the painting.[2] We use the case of rural electrification to show that even in a rich literature the pointillist painting is hard to compile and the dots on the canvas leave a lot of room for interpretation. We furthermore argue that in highly contested policy areas, even well-meaning policymakers will use this wiggle room to pursue their interests. Next, we argue that the practice of design-based research, despite its intellectual beauty to identify causality, is not immune to other biases stemming from questionable research practices, underpowered designs, overgeneralization, and publication bias. This further complicates the use of evidence in the policy landscape.

To conclude, we believe this observation is not particular to electrification. For most interventions, contexts are too heterogeneous to expect a coherent evidence picture to emerge within a timeframe that is helpful for much more urgent policy decisions. As a way forward for rural electrification and other development policies, we propose a humbler approach that focusses on improving the intervention under evaluation and refrains from undue generalization. Moreover, beyond development policy, more research is needed to study and improve the functioning of the science-policy interface.

2. The credibility revolution in the electrification literature

Prior to the credibility revolution, empirical research on rural electrification had been conducted for many decades and had recurrently featured insightful studies based on various methods. Nonetheless, it is a showcase example for what the credibility revolution rightly criticized in the 2000s: many studies made some sort of causal inference based on a simple comparison of people or regions with and without access to electricity (see Peters 2009).[3] That has changed over the past 15 years or so, with an increasing number of published studies that reveal more sensitivity for the problems of selection bias. The methodological portfolio first covered quasi-experimental matching and difference-in-difference designs, but increasingly also IVs.

[2] For the sake of completeness, in this allegory Duflo (2020) refers to RCTs alone. Most proponents of the "experimentalist paradigm" would extend this epistemology to other non-randomized design-based methods like IVs, regression discontinuity design and difference-in-differences; see for example Angrist and Pischke (2010) for a brief reference to this epistemology. Yet, clear statements and instructions on how the evidence is supposed to be compiled are very rare, in both textbooks and declaration-like papers.

[3] For more general cases beyond the electrification example, see Frondel and Schmidt (2005), Ravallion (2007) and Schmidt (2001).

Especially the latter method is the basis of several influential papers published in the profession's leading journals, most notably Dinkelman (2011) and Lipscomb et al. (2013). The decentralization of electricity access also facilitated randomization, so that the first RCTs appeared in the mid-2010s (Aklin et al. 2017; Furukawa 2014; Grimm et al. 2017). RCTs for power infrastructure in most settings proved to be infeasible for political or budgetary reasons, but in collaboration with the Kenyan utility, scholars even managed to randomize grid access (Lee et al. 2020a). Yet, for on-grid electrification, quasi-experimental methods and especially IVs continue to be the dominant identification strategies, while for off-grid solar several RCTs exist.

This wave of intense design-based research was followed by a battery of overview papers, systematic reviews and meta-analyses (henceforth 'reviews'; Bayer et al. 2020; Bernard 2012 ; Blimpo and Cosgrove-Davies 2019; Bonan et al. 2017; Bos et al. 2018; Hamburger et al. 2019; Jeuland et al. 2021 ; Jimenez 2017; Lee et al. 2020b; Morrissey 2018; Perdana et al. 2020; Peters and Sievert 2016).[4] The research community has hence not only generated the dots on Duflo's pointillist painting but also invested in compiling what the painting shows. All these reviews diagnose a divide in the literature, that is, one set of studies comes to very positive conclusions about the development effects of electrification while another set of studies rather observes small or no effects.

To understand the policy implication of this, the size of the effect must be assessed in relationship to the costs. Here, it is important to distinguish between on-grid and off-grid electrification. Given the high cost of grid-based rural electrification, large positive effects are required to make the intervention cost-effective and even positive, while modest effects would advocate *against* the investment. Based on their finding of muted effects, Lee et al. (2020a) conclude that the investment into grid extension entails a "social surplus loss". For off-grid electrification such as small-scale solar, in contrast, even modest effects can render a cost-benefit analysis positive and suggest that promoting this technology is cost-effective – because of the considerably lower investment cost (Grimm et al. 2020).

The reviews refer to several potential explanations of the divide, but in our reading, two narratives stand out:[5] a regional divide and a methodological divide. Jeuland et al. (2021) is an insightful starting point. It does not delve into a narrative for the divide in the literature. Its main purpose, rather, is to comprehensively take stock of the literature. Jeuland et al. (2021) thereby illustrate how vast the evidence base is when a review is very inclusive. By covering

[4] While some of these reviews use a systematic and replicable inventory of the literature as it is done in classical meta-analyses, others are based on purposeful selections of influential papers. There is no study to our knowledge that uses meta-analytical methods to combine effects of several studies to one effect.

[5] As a matter of course, various other sources of heterogeneity might explain why effects of electrification differ across studies, for example differences in grid reliability (Allcott et al. 2016; Chakravorty et al. 2014), exposure to exogenous economic development (Fetter and Usmani 2024), size of the targeted communities (Burlig and Preonas 2024), complementary services like access to finance, and the duration of exposure to electricity (Nag and Stern 2023). Traces of these factors can be found throughout the reviews, but in our reading none of the reviews puts emphasis on them.

a generous list of journals as well as the grey literature, it shows that the electrification literature comprises some 2,000 studies. As an extreme case, one can draw from this large pool to compile the pointillist painting, even if there is certainly a broad consensus that many of the dots should be dismissed, for example because a study does not apply design-based methods that meet the credibility revolution's standards. All other reviews employ much more exclusive selections of the literature and most include design-based studies only.

The *regional narrative* for the divide in the literature points to the different development potentials in different regions and target populations (see, for example, Hamburger et al. 2019, Lee et al. 2020b, Peters and Sievert 2016). Hamburger et al. (2019) reveal that large parts of the design-based electrification literature are concentrated in just a few countries. Especially Sub-Saharan Africa is largely ignored. Related to this, Peters and Sievert (2016) argue that the large effects observed in some Latin American and Asian countries cannot be generalized to Sub-Saharan Africa because of different economic conditions at baseline. They also provide evidence for small effects from several Sub-Saharan Africa countries, which contrast with the much larger effects in the pre-existing literature. In a similar vein, Lee et al. (2020b) emphasize that, historically, electrification in most industrialized countries happened while the economies were on a growth trajectory. Evidence from such contexts is hence not transferable to places today where remote and typically underdeveloped areas are connected that are barely integrated in economic development processes.

The *methodological narrative* is raised mainly in Bayer et al. (2019) and Lee et al. (2020b). Bayer et al. (2019) establish that studies using randomized designs typically deliver smaller effects than those using quasi-experimental designs. They explain this by the selection bias inherent to non-randomized methods that inflates impact estimates. The pattern in their data is indeed striking, but an important caveat is that with one exception all RCTs were done on off-grid electrification technologies, not the grid. Grid extension programs are mostly evaluated using IVs as well as regression discontinuity and difference-in-difference designs. Lee et al. (2020b), focusing on grid electrification, point to the large number of IVs in that literature and suggest that "it is hard to rule out the possibility that the correlation between the instrument and the dependent variable runs through additional channels beyond electrification".

In fact, the heavy reliance on observational data and especially IVs is conspicuous in the electrification literature, and it might import influential risks of bias. Above all, the geographic IVs that are often used in electrification evaluations such as the land gradient or water flow are suspected of violating exclusion restrictions because they affect the causal network through many pathways, not just through electrification, the instrumented variable (Bensch et al. 2020; Gallen and Raymond 2023; Haveresch et al. 2024; Lal et al. 2024; Lee et al. 2020b; Mellon 2024). Another reason to be concerned is that these geographical IVs are often weak IVs, which is not a problem per se if appropriate remedies are used. But these remedies are less effective if weakness concurs with violated exclusion restrictions (Bensch et al. 2020) and if scholars screen specifications based on first-stage strength (Ankel-Peters et al. 2023a). Related to the screening aspect, IVs are suspected of being more prone to publication bias and

p-hacking (Brodeur et al. 2020), because "when using a non-experimental method like IV there are many points at which a researcher exercises discretion in ways that could affect statistical significance".[6]

Against this background, perhaps a new look at the divide in the literature is warranted: we contend that large effects are mainly found in studies using secondary data, especially those based on IVs. Much smaller effects are found in studies that use primary data to evaluate specific electrification programs, irrespective of whether they are RCTs or based on a difference-in-differences procedure. This covers studies like Lee et al. (2020a), an RCT, but also Bensch et al. (2019), Chaplin et al. (2017), Lenz et al. (2017) and Peters et al. (2011). We argue that the key difference is in the evaluation of a specific, well-identified intervention and the fact that own data was collected. This sharpens the understanding of the treatment, it increases the willingness to publish a null result, and thus it decreases incentives to "exercise discretion" (Brodeur et al. 2020) to find exciting positive effects. This is not to say that studies based on primary data are always better. They face obvious limitations like regional scope, and of course they are sometimes prone to cooptation by development agencies that fund both the policy intervention and the evaluation.

In any case, the electrification literature should be evaluated in the light of recent trends in the economics profession towards more transparency (Christensen and Miguel 2018). That requires sensitivity for pre-specification and robustness replicability as well as quantitative meta-analyses that account for potential publication bias (Andrews and Kasy 2019; Carter et al. 2019; Irsova et al. 2024) – something that has hitherto not been done.

3. Bayesian policymakers and reasoned intuition

The target audience of applied empirical research according to the evidence-based policy paradigm are policymakers[7]. Economists have started to examine the conditions under which policymakers indeed make use of available evidence (Banuri et al. 2019; Hjort et al. 2021; Vivalt et al. 2023). The underlying assumption often is that the evidence provides a scientifically clear picture. In practice, though, the evidence is often murky and contradictory, and subject to debates about methodological issues. The electrification literature is a showcase example for this. It is therefore important to ask how policymakers form their beliefs.

Ideally, policymakers and we, their academic advisors, are Bayesians: We have a prior which we update as new evidence comes in. The prior's responsiveness is a function of the evidence's methodological quality. That is, the prior is firmer and less responsive to new evidence the better the already existing evidence is. Likewise, it responds more to methodologically sound

[6] See as well Kranz and Pütz (2022) and Brodeur et al. (2022).

[7] We use this term broadly and include different actors at the science-policy interface, for example, decision makers in governmental agencies at the strategic and the operational level as well as policy advisory committees.

new evidence. This type of thinking, though, requires repeated appraisals of the incoming evidence. For this appraisal there exist no standards. At best, these appraisals are based on experience and expertise. In other words, we must use what Basu (2014) calls *reasoned intuition:* "intuition and gut feeling […] need to be held under the scanner of reason before we use them to translate experience and evidence into rules and behaviour and policy." Most policymakers have experience and expertise, so it is possible that reasoned intuition can work when policymakers come across new evidence.

Yet, so far, we have assumed benevolent policymakers while in practice they might have some sort of vested interests. This is, in many cases, not condemnable. For example, policymakers are typically civil servants and hence subscribe to a certain political agenda of the administration they represent. It is natural that policymakers extract from the evidence what serves their interest. A divided literature like the one on rural electrification provides the basis for confirmation bias as it is empirically diagnosed by Banuri et al. (2019). In a similar vein, Vivalt and Coville (2023) provide empirical evidence for what they call "asymmetric optimism": policymakers update more on good news than on bad news. Underpinning both studies is work with staff members of development agencies and, thus, policymakers similar to those who are responsible for rural electrification decisions.

Policymakers managing electrification portfolios can have agendas. For example, major development banks have a long history of investing in large infrastructure through grants and lending, and it is understandable that they – or some of their staff members – prefer on-grid electrification over off-grid electrification. Confirmation bias and asymmetric optimism might tempt them to seize that part of the literature that suggests substantial development effects of grid extension programs. Staff of solar advocacy organizations or private sector representatives seeking subsidies for their off-grid solar programs might, by contrast, prefer evidence suggesting only modest impacts of on-grid electrification. This would strengthen the cost-effectiveness of off-grid technologies. The hawker's tray of the electrification literature has much evidence to offer for both camps.

An informed debate between these two camps based on reasoned intuition is hence problematic. An additional important layer of complexity is that applying reasoned intuition is harder the more prevalent methodological concerns are that are not well understood within academia.[8] For example, academic debates do not converge when it comes to publication bias and how to account for it when making inference. Likewise, controversies about robustness in replications and reproductions are hard to settle among replicators and original authors (Ankel-Peters et al. 2023b; Ozier 2021). And while external validity is an accepted barrier between rigorous evidence and its policy relevance, the literature on how to account for it in the generalization of scientific results is nascent but so far inconclusive (Dehejia et al. 2019; Gechter 2023; Muller 2015, 2021; Peters et al. 2019; Pritchett and Sandefur 2015; Vivalt 2020).

[8] Vivalt and Coville (2023) also emphasize that potential biases of policymakers in reading the evidence is more problematic in the presence of biases in the underlying evidence, such as publication bias or lacking external validity.

Concerns about construct validity are less widely discussed and virtually absent in the economics literature, although they are of utmost importance for generalization across supposedly similar interventions (Esterling et al. 2023; Masselus et al. 2024; Pritchett et al. 2013). Such debates including their ambiguous outcomes are not a failure but rather a natural part of the scientific enterprise. Yet, they do pose major hurdles for the evidence-policy interface.

4. Conclusion and way forward

In this paper, we have argued that the evidence-based policy paradigm reaches its limits in the case of rural electrification. Policymakers with vested interests of different kinds will each find support for their respective agenda. But even benevolent policymakers might get into difficulties because of unresolved methodological debates in the literature. It is overly simplistic, though, to merely blame policymakers for extracting only a partial interpretation of the evidence. Academic researchers bear part of the responsibility in that they often communicate results with what Manski (2011, 2019) calls *incredible certitude*.

Manski stresses that the logic of any inference is: assumptions + data => conclusions. The rural electrification research community deserves to be applauded for the many systematic reviews it has produced, to which we owe the consolidated understanding that this literature is divided. Most individual papers, though, *wishfully extrapolate* (again, Manski) their data to much too strong conclusions, by imposing heavy assumptions. These assumptions are only partly made transparent and range from external validity concerns to a much weaker robustness than what is communicated in the papers.

The patterns we have diagnosed in this paper are not a peculiarity of rural electrification.[9] Many literatures, including those that have been subject to a myriad of design-based impact evaluations, exhibit fuzzy pointillist paintings and methodological issues related to external validity and reproducibility. What are the implications for the learning model in the electrification literature and beyond? One response would be to more and more design-based studies, accompanied by robustness replications ensuring that the right inference is being made, and hope for a clearer picture emerging in the literature soon. However, "the pace of politics is faster than the pace of scientific consensus formation" (Collins and Evans 2022, p. 241), that is, important policy decisions are due before this will happen.[10]

Theory-based evaluation might help to accelerate this process (Duflo et al. 2007; White 2009). Clearly outlined theory can identify mechanisms, which are then tested in (quasi-)ex-

[9] Problematic or controversial patterns in other literatures related to environmental economics and policy have been diagnosed, for example, in Ferraro and Shukla (2020; 2023), Vrolijk and Sato (2023), Bagilet and Zabrocki-Hallak (2022), and Whittington et al. (2012).

[10] David Card, in his Nobel Prize lecture in 2021, expressed his optimism that the debate around minimum wages that started in the early 1990s might converge to a common understanding in "another decade or two" (Card 2022).

periments. The hope is that such mechanisms are less context dependent and hence more generalizable than effects of the whole program, which is often a bundle of interventions (Ludwig et al. 2011). It is true that much of the literature on rural electrification is lacking such a clear theory, and mechanisms such as productive use take-up are rarely tested in a theory-grounded manner. This would require pre-specification of hypotheses, not just explorative heterogeneity analysis (which is indeed done in several studies). A clearly outlined theory would also render Manski's *wishful extrapolation* more difficult because the theoretical foundation would expose the assumptions underlying the extrapolation.

In the meantime, a pragmatic way-forward for design-based research is to reflect on a humbler epistemic paradigm: impact evaluations could focus on informing the specific program under evaluation only and widely refrain from generalization to other contexts.[11] Impact evaluations would then rather be a feature for internal program management than for global learning processes. Elements of this can be found also in proposals from within the credibility revolution movement (see Banerjee et al. 2017; Duflo 2017). This focus on improving a specific program might also reduce some of the pressures leading to publication bias and the use of questionable research practices. Yet the current reward system in academia and from funding agencies does probably not incentivize such a humbler approach.

More generally, a research program is needed in the economics profession on how the science-policy interface works and how it can be improved. Absent formal evidence clearing houses like the World Health Organization or the Intergovernmental Panel on Climate Change, policy often relies on scientific advisory boards or bilateral consultations to be backed up by scientific expertise. That is fine, but policymakers need to be sensitized to the pitfalls of evidence-based policy advice outlined in this paper. Ultimately, we need a better methodology for how to organize and synthesize knowledge formation in economics – a slightly belated version of "studies of expertise and experience" (Collins and Evans 2002). This will raise many important downstream questions for the economics profession, ushering in a veritable research program.

Competing interests:

One of the authors has made several contributions to the literature under review in this paper. Beyond this, the authors declare no competing interests.

[11] See Ferraro et al. (2023) for a related proposal for environmental programs.

References

Allcott, H., Collard-Wexler, A. and O'Connell, S.D., (2016). How do electricity shortages affect industry? Evidence from India. *American Economic Review*, *106*(3), pp.587-624.

Andrews, I. and Kasy, M. (2019). Identification of and correction for publication bias. *American Economic Review*, *109*(8), pp.2766-2794.

Angrist, J. D., & Pischke, J. S. (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.

Ankel-Peters, J., Bensch, G. and Vance, C. (2023a). *Spotlight on researcher decisions: Infrastructure evaluation, instrumental variables, and specification screening* (No. 991). Ruhr Economic Papers.

Ankel-Peters, J., Fiala, N. and Neubauer, F. (2023b). *Is economics self-correcting? Replications in the American Economic Review* (No. 1005). Ruhr Economic Papers.

Aklin, M., Bayer, P., Harish, S.P. and Urpelainen, J. (2017). Does basic energy access generate socioeconomic benefits? A field experiment with off-grid solar power in India. *Science advances*, *3*(5), p.e1602153.

Bagilet V. and Zabrocki-Hallak L. (2022). Why Some Acute Health Effects of Air Pollution Could Be Inflated. *Institute for Replication (I4R) Discussion Paper Series*.

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., … & Walton, M. (2017). From proof of concept to scalable policies: Challenges and solutions, with an application. *Journal of Economic Perspectives*, *31*(4), 73-102.

Banuri, S., Dercon, S., & Gauri, V. (2019). Biased policy professionals. *World Bank Economic Review*, *33*(2), 310-327.

Barnes, D.F., (2010). *The Challenge of Rural Electrification – Strategies for Developing Countries*. Routledge.

Barnes, D.F. and Binswanger, H.P., (1986). Impact of rural electrification and infrastructure on agricultural changes, 1966-1980. *Economic and Political Weekly*, pp.26-34.

Basu, K. (2014). Randomisation, causality and the role of reasoned intuition. *Oxford Development Studies*, *42*(4), pp.455-472.

Bayer, P., Kennedy, R., Yang, J., & Urpelainen, J. (2020). The need for impact evaluation in electricity access research. *Energy Policy*, 137, 111099.

Bensch, G., Gotz, G. and Ankel-Peters, J. (2020). Effects of rural electrification on employment: A comment on Dinkelman (2011). https://osf.io/preprints/metaarxiv/zhn9b

Bensch, G., Cornelissen, W., Peters, J., Wagner, N., Reichert, J. and Stepanikova, V. (2019). Electrifying Rural Tanzania. A Grid Extension and Reliability Improvement Intervention. https://www.econstor.eu/handle/10419/222259

Bernard, T. (2012). Impact analysis of rural electrification projects in sub-Saharan Africa. *The World Bank Research Observer*, *27*(1), 33-51.

Biddle, J.E. and Hamermesh, D.S. (2017). Theory and measurement: Emergence, consolidation, and erosion of a consensus. *History of Political Economy*, *49*(Supplement), pp.34-57.

Blimpo, M. P., & Cosgrove-Davies, M. (2019). Electricity access in Sub-Saharan Africa: Uptake, reliability, and complementary factors for economic impact. World Bank Publications.

Bonan, J., Pareglio, S., & Tavoni, M. (2017). Access to modern energy: a review of barriers, drivers and impacts. *Environment and Development Economics*, *22*(5), 491-516.

Bos, K., Chaplin, D., & Mamun, A. (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.

Brodeur, A., Cook, N. and Heyes, A., 2020. Methods matter: P-hacking and publication bias in causal analysis in economics. *American Economic Review*, *110*(11), pp.3634-3660.

Brodeur, A., Cook, N. and Heyes, A., 2022. Methods matter: p-hacking and publication bias in causal analysis in economics: reply. *American Economic Review*, *112*(9), pp.3137-3139.

Burlig, F. and Preonas, L., (2024). Out of the darkness and into the light? Development effects of rural electrification. *Journal of Political Economy*, forthcoming.

Card, D., 2022. Design-based research in empirical microeconomics. *American Economic Review*, *112*(6), pp.1773-1781.

Carter, E.C., Schönbrodt, F.D., Gervais, W.M. and Hilgard, J. (2019). Correcting for bias in psychology: A comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science*, *2*(2), pp.115-144.

Chaplin, D., Mamun, A., Protik, A., Schurrer, J., Vohra, D., Bos, K., Burak, H., Meyer, L., Dumitrescu, A., Ksoll, C. and Cook, T. (2017). Grid electricity expansion in Tanzania by MCC: Findings from a rigorous impact evaluation. *Report Submitted to the Millennium Challenge Corporation. Washington, DC: Mathematica Policy Research*, pp.6-13.

Chakravorty, U., Pelli, M. and Marchand, B.U., (2014). Does the quality of electricity matter? Evidence from rural India. *Journal of Economic Behavior & Organization*, *107*, pp.228-247.

Christensen, G. and Miguel, E. (2018). Transparency, reproducibility, and the credibility of economics research. *Journal of Economic Literature*, *56*(3), pp.920-980.

Collins, H.M. and Evans, R., (2002). The third wave of science studies: Studies of expertise and experience. *Social Studies of Science*, *32*(2), pp.235-296.

Deaton, A., (2020). *Randomization in the tropics revisited: a theme and eleven variations*. National Bureau of Economic Research Working Paper Series, No. W27600.

Deaton, A. and Cartwright, N., (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, *210*, pp.2-21.

Dehejia, R., Pop-Eleches, C., & Samii, C. (2019). From local to global: External validity in a fertility natural experiment. *Journal of Business & Economic Statistics*, 1-27.

Devine, W.D., (1983). From shafts to wires: Historical perspective on electrification. *The Journal of Economic History*, *43*(2), pp.347-372.

Dinkelman, T. (2011). The effects of rural electrification on employment: New evidence from South Africa. *American Economic Review*, *101*(7), pp.3078-3108.

Drèze, J. (2020). Policy beyond evidence. *World Development*, *127*, 104797.

Duflo, E. (2004). Scaling up and evaluation. In *Annual World Bank Conference on Development Economics 2004* (pp. 341-369).

Duflo, E. (2017). The economist as plumber. *American Economic Review*, *107*(5), pp.1-26.

Duflo, E. (2020). Field experiments and the practice of policy. *American Economic Review*, *110*(7), pp.1952-1973.

Duflo, E., Glennerster, R. and Kremer, M., 2007. Using randomization in development economics research: A toolkit. *Handbook of Development Economics*, *4*, pp.3895-3962.

Esterling, K.M., Brady, D. and Schwitzgebel, E. (2023). *The necessity of construct and external validity for generalized causal claims* (No. 18). I4R Discussion Paper Series.

Ferraro, P. J., & Shukla, P. (2020). Is a replicability crisis on the horizon for environmental and resource economics. *Review of Environmental Economics and Policy*, 14, issue 2, pp. 339–351

Ferraro, P.J. and Shukla, P. (2023). Credibility crisis in agricultural economics. *Applied Economic Perspectives and Policy*, *45*(3), pp.1275-1291.

Ferraro, P.J., Cherry, T.L., Shogren, J.F., Vossler, C.A., Cason, T.N., Flint, H.B., Hochard, J.P., Johansson-Stenman, O., Martinsson, P., Murphy, J.J. and Newbold, S.C., (2023). Create a culture of experiments in environmental programs. *Science*, *381*(6659), pp.735-737.

Fetter, T.R. and Usmani, F., (2024). Fracking, farmers, and rural electrification in India. *Journal of Development Economics*, p.103308.

Frondel, M. and Schmidt, C.M., (2005). Evaluating environmental programs: The perspective of modern evaluation research. *Ecological Economics*, *55*(4), pp.515-526.

Furukawa, C. (2014). Do solar lamps help children study? Contrary evidence from a pilot study in Uganda. *Journal of Development Studies*, *50*(2), pp.319-341.

Gechter, M. (2023). Generalizing the Results from Social Experiments: Theory and Evidence from India. *Journal of Business & Economic Statistics*, forthcoming.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, *5*(10), 3-8.

Gallen, T. and Raymond, B. (2023). Broken Instruments.

Grimm, M., Munyehirwe, A., Peters, J. and Sievert, M. (2017). A first step up the energy ladder? Low-cost solar kits and household's welfare in rural Rwanda. *The World Bank Economic Review*, *31*(3), pp.631-649.

Grimm, M., Lenz, L., Peters, J. and Sievert, M. (2020). Demand for off-grid solar electricity: Experimental evidence from Rwanda. *Journal of the Association of Environmental and Resource Economists*, *7*(3), pp.417-454.

Hamburger, D., Jaeger, J., Bayer, P., Kennedy, R., Yang, J., & Urpelainen, J. (2019). Shades of darkness or light? A systematic review of geographic bias in impact evaluations of electricity access. *Energy Research & Social Science*, 58, 101236.

Haveresch, N., Ankel-Peters, J. and Bensch, G. 2023. A slippery slope: topographic variation as an instrumental variable. Mimeo.

Heckman, J.J. and Urzua, S. (2010). Comparing IV with structural models: What simple IV can and cannot identify. *Journal of Econometrics*, *156*(1), pp.27-37.

Hjort, J., Moreira, D., Rao, G. and Santini, J.F. (2021). How research affects policy: Experimental evidence from 2,150 brazilian municipalities. *American Economic Review*, *111*(5), pp.1442-1480.

Irsova, Z., Doucouliagos, H., Havranek, T. and Stanley, T.D., (2024). Meta-analysis of social science research: A practitioner's guide. *Journal of Economic Surveys,* forthcoming.

Jeuland, M., Fetter, T.R., Li, Y., Pattanayak, S.K., Usmani, F., Bluffstone, R.A., Chávez, C., Girardeau, H., Hassen, S., Jagger, P. and Jaime, M.M., […] (2021). Is energy the golden thread? A systematic review of the impacts of modern and traditional energy use in low-and middle-income countries. *Renewable and Sustainable Energy Reviews*, *135*, p.110406.

Jimenez, R. (2017). Development effects of rural electrification. Inter-American Development Bank.

Kranz, S. and Pütz, P., 2022. Methods matter: P-hacking and publication bias in causal analysis in economics: Comment. *American Economic Review*, *112*(9), pp.3124-3136.

Lal, A., Lockhart, M., Xu, Y. and Zu, Z., 2024. How much should we trust Instrumental Variable estimates in Political Science? Practical advice based on 67 replicated Studies. *Political Analysis*, forthcoming.

Lee, K., Miguel, E. and Wolfram, C., (2020a). Experimental evidence on the economics of rural electrification. *Journal of Political Economy*, *128*(4), pp.1523-1565.

Lee, K., Miguel, E., & Wolfram, C. (2020b). Does Household Electrification Supercharge Economic Development? *Journal of Economic Perspectives*, 34(1), 122-44.

Levin, T. and Thomas, V.M. (2016). Can developing countries leapfrog the centralized electrification paradigm? *Energy for Sustainable Development*, *31*, pp.97-107.

Lenz, L., Munyehirwe, A., Peters, J. and Sievert, M. (2017). Does large-scale infrastructure investment alleviate poverty? Impacts of Rwanda's electricity access roll-out program. *World Development*, *89*, pp.88-110.

Lipscomb, M., Mobarak, A.M. and Barham, T. (2013). Development effects of electrification: Evidence from the topographic placement of hydropower plants in Brazil. *American Economic Journal: Applied Economics*, *5*(2), pp.200-231.

Ludwig, J., Kling, J.R. and Mullainathan, S., 2011. Mechanism experiments and policy evaluations. *Journal of Economic Perspectives*, *25*(3), pp.17-38.

Manski, C. F. (2011). Policy analysis with incredible certitude. *Economic Journal*, *121*(554), F261-F289.

Manski, C.F. (2019). Communicating uncertainty in policy analysis. *Proceedings of the National Academy of Sciences*, *116*(16), pp.7634-7641.

Masselus, L., Ankel-Peters, J. and Petrik, C. (2023): Lost in the Design Space? Construct Validity in the Microfinance Literature. Mimeo.

Mellon, J. (2024). Rain, Rain, Go Away: 195 Potential Exclusion-Restriction Violations for Studies Using Weather as an Instrumental Variable. *American Journal of Political Science*, forthcoming.

Morrissey, James (2018). Linking Electrification and Productive Use. Oxfam Research Backgrounder Series.

Muller, S.M., (2015). Causal interaction and external validity: Obstacles to the policy relevance of randomized evaluations. *The World Bank Economic Review*, *29* (suppl_1), pp.S217-S225.

Muller, S.M., (2021). Randomised trials in economics. In *A modern guide to philosophy of economics* (pp. 90-126). Edward Elgar Publishing.

Muller, S.M., (2023). Is economics credible? A critical appraisal of three examples from microeconomics. *Journal of Economic Methodology*, *30*(2), pp.157-175.

Nag, S. and Stern, D.I., 2023. Are the benefits of electrification realized only in the long run? Evidence from rural India. Evidence from Rural India. https://ssrn.com/abstract=4591072

Ozier, O. (2021). Replication Redux: The reproducibility crisis and the case of deworming. *The World Bank Research Observer*, *36*(1), pp.101-130.

Panhans, M.T. and Singleton, J.D. (2017). The empirical economist's toolkit: from models to methods. *History of Political Economy*, *49*(Supplement), pp.127-157.

Perdana, A., D. Glandon, N. Moore and B. Snilsveit (2020) How do electricity access interventions affect social outcomes? A forthcoming systematic review. International Initiative for Impact Evaluation (3ie).

Peters, J., & Sievert, M. (2016). Impacts of rural electrification revisited–the African context. *Journal of Development Effectiveness*, 8(3), 327-345.

Peters, J. (2009). Evaluating Rural Electrification Projects-Methodological Approaches. *Ruhr Economic Papers*Rio.

Peters, J., Vance, C. and Harsdorff, M. (2011). Grid extension in rural Benin: Micro-manufacturers and the electrification trap. *World Development*, *39*(5), pp.773-783.

Pritchett, L. and Sandefur, J. (2015). Learning from experiments when context matters. *American Economic Review*, *105*(5), pp.471-475.

Pritchett, L., Samji, S. and Hammer, J.S. (2013). It's all about MeE: Using Structured Experiential Learning ('e') to crawl the design space. *Center for Global Development Working Paper*, (322).

Ravallion, M., (2007). Evaluating anti-poverty programs. *Handbook of Development Economics*, *4*, pp.3787-3846.

Ravallion, M. (2009). Should the randomistas rule? *The Economists' Voice*, *6*(2).

Ravallion, M. (2020). Should the Randomistas (Continue to) Rule? *Randomized Control Trials in the Field of Development: A Critical Perspective*.

Rodrik, D. (2008). The new development economics: we shall experiment, but how shall we learn?

Rose, J.K. (1940). Rural electrification: A field for social research. *Rural Sociology*, *5*(4), pp.411-26.

Schmidt, C.M., (2001). Knowing what works: the case for rigorous program evaluation. *Available at SSRN 273173*.

Vivalt, E. (2020). How much can we generalize from impact evaluations? *Journal of the European Economic Association*, *18*(6), pp.3045-3089.

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

Vrolijk, K. and Sato, M. (2023). Quasi-experimental evidence on carbon pricing. *The World Bank Research Observer*, p.lkad001.

White, H., 2009. Theory-based impact evaluation: principles and practice. *Journal of Development Effectiveness*, *1*(3), pp.271-284.

Whittington, D., Jeuland, M., Barker, K. and Yuen, Y., 2012. Setting priorities, targeting subsidies among water, sanitation, and preventive health interventions in developing countries. *World Development*, *40*(8), pp.1546-1568.

World Bank (2018). Africa's Pulse, Spring 2018: Analysis of Issues Shaping Africa's Economic Future (April), World Bank, Washington, DC. Doi: 10.1596/978-1-4648-1291-0

Young, K., Ashby, D., Boaz, A. and Grayson, L., 2002. Social science and the evidence-based policy movement. *Social Policy and Society*, *1*(3), pp.215-224.