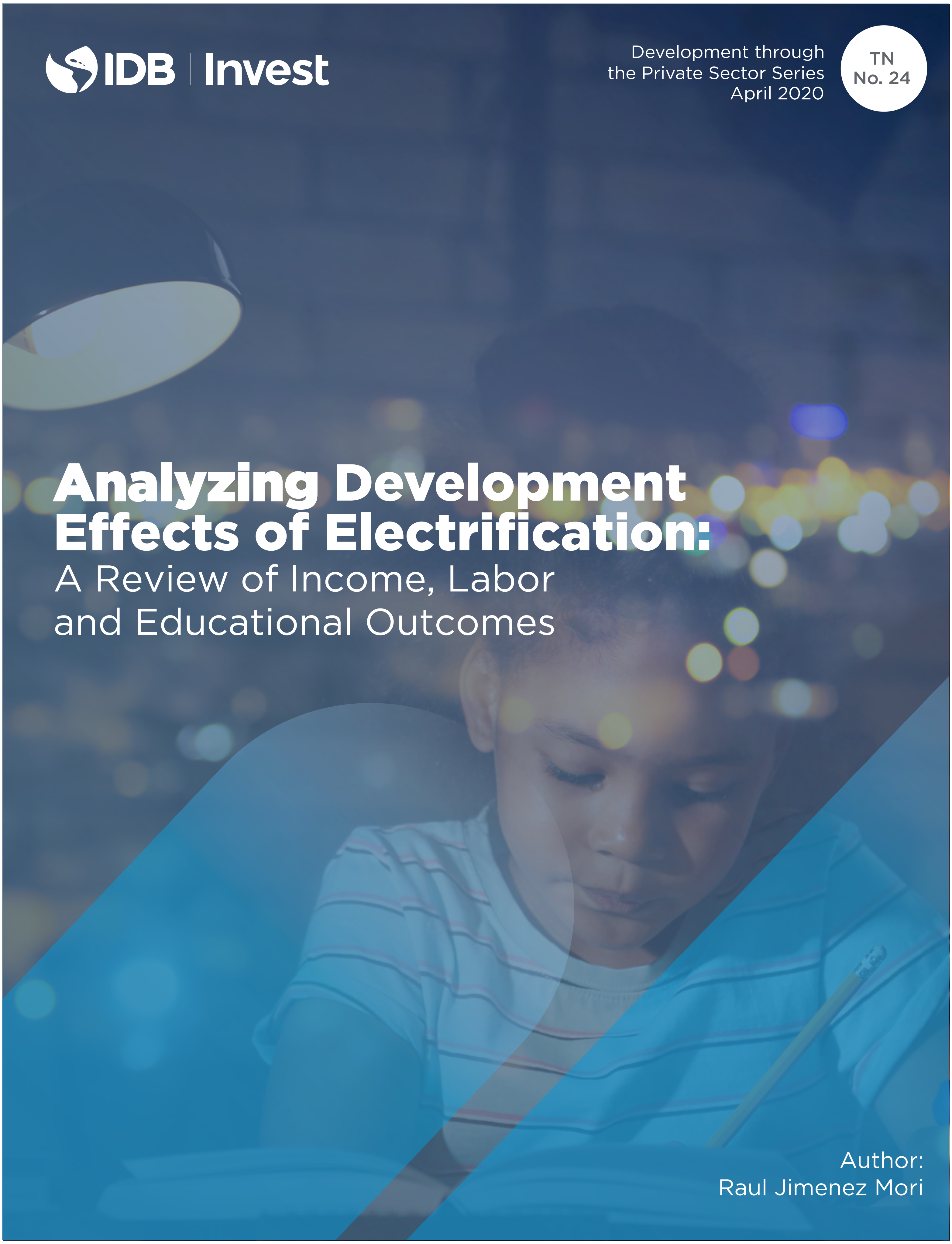


IDB Invest

Development through
the Private Sector Series
April 2020

TN
No. 24

Analyzing Development Effects of Electrification:
A Review of Income, Labor and Educational Outcomes

Author:
Raul Jimenez Mori

IDB Invest

Analyzing Development Effects of Electrification:

A Review of Income, Labor and Educational Outcomes

Cover page design: Jimena Vazquez

April 2020

Analyzing Development Effects of Electrification: A Review of Income, Labor and Educational Outcomes

Raul Jimenez Mori*

Abstract

Globally, substantial public and private investment has been allocated to electrification, the benefits of which have remained largely, and intuitively, undisputed. However, a close look at the literature suggests that those benefits are far from consistent across different contexts. In fact, estimates of the impacts of electrification are widely heterogeneous and inconclusive. For example, almost 40 percent of the literature's estimates find no significant effects on income, with similar results for labor and educational outcomes. This paper synthesizes the impact evaluation studies published between mid-1980s and mid-2019, focusing on the effects of electrification on education, labor and income outcome-categories. Those studies indicate electrification effects (and 95 percent confidence intervals (CI)) of around eight percent in school enrollment (CI=0.04-0.13), 17 percent in employment (CI=0.02-0.34), and 24 percent in income (CI= 0.15-0.35). This paper shows, however, that those averages are biased upward, suggesting the presence of selective publication. The regression estimates indicate a mean-corrected, statistically significant effect of 2.6 percent for school enrollment (CI= 0.01-0.04), 5.3 percent for employment (0.04-0.07), and seven percent for income (0.05-0.09). The study design helps explain the estimates' heterogeneity—i.e., methodologies and data structures requiring less stringent assumptions tend to return lower estimated effects. These findings suggest that the observed heterogeneity reflects actual dispersion in the underlying effects associated to gender differences, as well as, contextual characteristics of the case study, such as country income per-capita and length of the evaluation period.

Keywords: Electrification, Income, Schooling, Employment, Publication bias, Meta-analysis.
JEL Codes: O13, Q40, C80.

*Contact: rjimenez@iadb.org. I thank Andrea Guerrero, Emily Brearley and Diego Margot for their helpful comments. All remaining errors are the author's responsibility. The views expressed here are my own and not necessarily those of the Inter-American Development Bank Group. An online appendix providing the data and the Stata code used in this paper is openly available at http://doi.org/[doi](pending).

1 Introduction

While the expansion of electricity grid infrastructure is, to a great extent, a demand-driven process, electrification programs are also typically supported by society as a whole, under the widely accepted premise that access to electricity will make households better off, and increase productivity. This long-standing consensus that electrification is a development milestone is reflected, for example, in the Sustainable Development Goal of achieving universal access to affordable, reliable and modern energy services by 2030. In support of this goal, public and private initiatives have actively mobilized financial and technical resources, making electricity available to millions of people.[1]

It is surprising, however, that the evidence presented by the literature on the impacts of electrification programs on welfare outcomes varies so widely. For example, almost 40 percent of the literature's estimates find no statistically significant effects on income, with similar results for the labor and educational dimensions. Those results seem to conflict with the recent emphasis that policy makers and donors have placed on such programs. A critical concern is that this emphasis may indicate the presence of over expectations that, in turn, could undermine the design of adequate programs and policies (e.g., by overshooting expected results). Therefore, it is necessary to have a clear perspective on the effectiveness of electrification programs and to understand why their effectiveness may vary.

This paper presents a synthesis of the literature evaluating the impacts of access to, and improvements in, electricity services, with a focus on education, labor, and income outcomes. The analysis is based on mixed-effects multilevel regression, which allows consideration of the presence of selective publication; as well as examining the role of studies characteristics in the heterogeneity of the estimates. Also, the analysis provides benchmark estimates of the underlying effects of electrification.

The analysis is based on a unique data set composed of 196 estimates from 49 impact evaluation studies published between 1983 and mid-2019. The results from impact evaluations are rarely directly comparable, due to different specifications, outcomes metrics, populations and/or evaluation periods. Therefore, the estimated effects are expressed as a percentage change relative to the control mean. Consistent with this approach, and with the data requirements of meta-analysis methods, the examination uses only estimates for which the control means, and standard errors (or t-statistics) are reported in order to analyze their precision. In addition, the data set contains a rich set of the studies' characteristics, including: type of treatment (continuous or discrete); type of data and identification strategy; type of publication (journal or working paper); area of analysis (urban, rural or both); evaluation period; country; and whether the outcome is gender-specific.

[1]According to IEA [2018], as of 2018, over 550 million people have gained access since 2011, and it will required over US$660 billion to reach universal access by 2030.

The results shows a substantial heterogeneity in the estimated effects of electrification and sizeable differences between them, and the bias-corrected mean effects in all outcome categories. Overall, the literature reviewed suggests that electrification leads to weighted average (median) increases of around six percent (five percent) in school enrollment, 13 percent (ten percent) in employment, and 23 percent (16 percent) in income. However, more than one half and one third of the estimates find no statistically significant effects in the cases of enrollment and schooling, respectively. Similarly, around half of the estimates indicate non-significant impacts on labor and income outcomes. Accounting for this variability, the average reported estimates are equivalent to less than a 0.2 standard deviation, a level considered low by the standards of the meta-analysis literature. Consistent with this, the publication-bias-corrected means suggest smaller effects, in the order of three percent, 5.2 percent and seven percent, respectively. This examination suggests that the observed heterogeneity reflects true variation in the underlying effects among different populations, while sampling error plays a lesser role. In addition, the augmented meta-regressions provide greater insights into the role of study characteristics in explaining the variation in the observed estimates. Specifically, more-rigorous methodological approaches and richer data structures seem to lead to more-conservative estimates in all outcome categories. Although case-study characteristics have a less clear association with the direction and statistical significance of the estimates, they suggest that a country's level of economic development may play a role in how electricity delivers the expected benefits. There are also indications that electrification returns gender-differentiated effects, greatly benefiting women and girls.

This paper joins the growing body of meta-analysis literature addressing estimates' heterogeneity based on a rich set of covariates. To the best of my knowledge, this is the first systematic review on the impacts of electrification, quantifying the uncertainty in their estimates, examining the presence of selective publication and estimating the bias-corrected underlying effects. The paper's main contribution is twofold: 1) to help in the formation of realistic expectations regarding electrification programs; and 2) to provide insights into the conflicting estimates that may inform public policy. While this analysis finds that the effects of electrification seem to be widely heterogeneous and relatively small, on average, the meta-regressions suggest that those effects maintain statistical and economic significance. Electrification should not be seen as a silver bullet, but should be viewed in context, as a single dimension of the development toolbox that policy makers can use under a multi-sectoral approach.

Section 2 briefly outlines how access to improved electricity services is expected to translate into social welfare gains. Section 3 presents the data set used in this paper, along with the procedures used to systematize the estimations from the different studies. Section 4 describes the methodology used to investigate the factors driving the heterogeneous results in the literature. Section 5 presents the findings, and Section 6 concludes.

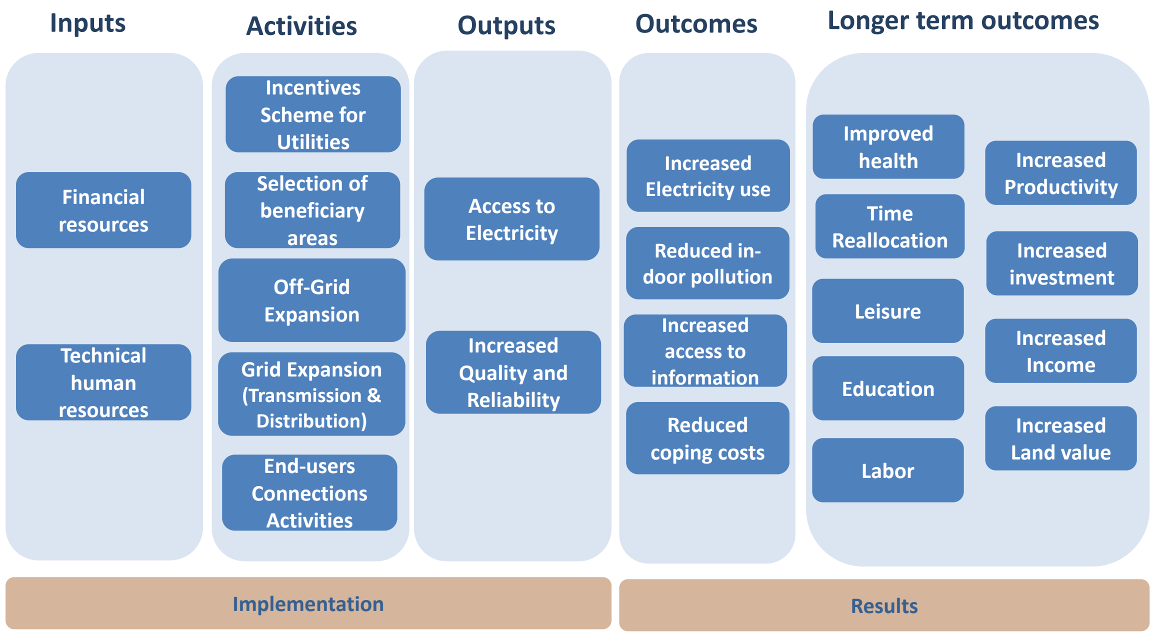
2 Theory of Change from Electricity to Development

This section presents a conceptual framework of the causal links typically addressed in the empirical literature of electrification. Figure 1 presents a simplified causal chain from inputs, to the expected impacts of electrification which can partially frame the literature on electrification.[2] This causal grid can be broadly separated into an implementation side, and a results side. The former groups the inputs (including financial and human resources) and the activities (such as grid expansion, installation of isolated systems, implementation of connection-fee subsidies, and/or social tariff programs) aimed at delivering the target output. The output can be divided into new connections and increased quality of the services.

Regarding the results side, at the household level, the literature suggests three main outcomes: decreased indoor pollution, increased electricity use, and changes in time allocation. First, it is argued that access to electricity reduces the consumption of low-quality, dirty fuels for lighting (such as kerosene or candles), thereby reducing emissions of polluting gases in the home. This would, in turn, produce health improvements, mainly regarding respiratory illnesses, especially among household members who spend more time inside the dwelling (women and children). Second, new access to electricity, in addition to offering a more efficient energy source, could also reduce coping costs and energy expenditures, thus increasing disposable income at the household level. Third, access to and use of electricity also implies an extension of usable hours during the night, potentially allowing for increased leisure and/or productive activities (educational or labor-related).

[2]The figure does not present some outcomes such as safety perceptions, crime, fertility, etc., though it should be noted that recent research has shown that electrification may have an impact on those outcomes

Figure 1: Causal Grid from Electricity to Development



The potential intermediate effects on health, labor, and education are expected to improve the quality of human capital, leading to income gains over the medium and long run. These income gains occur as a result of the more productive use of current capacities or of gains in human and physical capital. Over time, the electricity infrastructure would have external effects on variables such as land and household value, labor opportunities, and productive investments in the electrified areas [e.g. Grogan [2018], Barron and Torero [2017], Lipscomb et al. [2013]].

The quality and reliability of electricity services operate via similar channels, with the same long-term impacts. That is, having low-quality or unreliable service limits the extent to which households can take full advantage of the available infrastructure. In the case of firms, the availability of electricity infrastructure and changes in its quality, effect the patterns of electricity use and related production factors. Systematic outages may severely limit the extent to which firms can take full advantage of physical and human factors of production, reducing overall productivity [e.g. Fisher-Vanden et al. [2015], Rud [2012]].

A key point is that the realization and magnitude of the long-term effects of electrification, depend on behavioral considerations at the household and firm levels. In the case of households, time allocation among family members, education, and leisure plays a crucial role in defining impacts in the medium and long term. For firms, investment decisions may depend on the quality and reliability of electricity, as well as on the costs of coping with energy scarcity. The incorporation of nuanced responses to these key factors within the program design, and their corresponding economic analyses, could improve the effectiveness of electricity investments.

In fact, many of the appealing policy features of impact evaluations—such as their ability to facilitate transparency and accountability, to identify best practices, and to determine policy alternatives—are severely restricted if evaluations do not start in the early stages of program development, engaging in the details of policy making [Torero, 2015, Gertler et al., 2016, Duflo, 2017]. Indeed, impact evaluation plays a relevant role not only in quantifying which interventions work, but maybe most importantly in tracing the factors that influence their effectiveness and efficiency. For example, identifying quality and reliability of actual service provided may explain the extent to which new infrastructure is actually capable of impacting user decisions. Similarly, deficiencies in the implementation side may lead to problems such as cost overruns, that would effect the cost-benefit balance of the intervention. Despite the relevance of the implementation phase, as well as understanding the causal channels; an apparent limitation of the literature is that most impact evaluations, especially ex post evaluations, start later and are limited mainly to the results side, reducing their capacity to inform the design and implementation of programs.

Therefore, though the previous causal grid may frame the direction of the expected effects, important deviations may occur due to differences and limitations of the study which can blur our understanding of why the intervention worked or not. On the other hand, the size of the impact may also depend on context specific factors. For example, the marginal effect of electrification may be greater in higher income areas or countries, with availability of other complementary infrastructure, or better access to finance. In other terms, external validity across case studies may need to be considered before extrapolating results across different contexts.

3 Data

This section describes the procedure followed to collect and systematize relevant information from different case studies. The *inclusion criteria* were restricted to studies that explicitly attempted to estimate the causal effects of access to, or improvements in, electricity services over income, labor and educational outcomes categories. This includes studies on firms and households. Due to the heterogeneity of the indicators, similar dependent variables are grouped within these three main categories. In the case of income, household income or expenditures, household per-capita income at a monthly or a yearly frequency are considered to be equivalent. Similarly, for studies focus on firms, revenues and profits are considered to reflect the income dimension. The labor category groups hours worked and employment status and applies for studies addressing firms and households. The education category considers enrollment, years of schooling and time allocated to study.

Other relevant impacts are not included because of the scarcity of estimates for equivalent indicators. Therefore, dimensions not accounted in this analysis are, for example, health, firm productivity, safety, fertility, energy expenditures,fuel choice, etc. Similarly, this paper does not address spill-over effects of electrification.

The *data collection* began by reviewing major databases and search engines were consulted (i.e. EBSCO, Proquest, ScienceDirect, google, and google scholar), using a variety of search terms (i.e. electrification, impacts, effects, electricity access, electricity quality, economic, development). The results included various types of interventions, such as on- and off-grid programs of urban and/or rural scope; with public, private or mixed executing agencies. The data set also includes studies focusing on different units of observation (i.e., firms, households), under any level of aggregation (e.g., household, municipality, village-level). Published articles, working papers, reports and book chapters are included. The collection of papers concluded in June 2019, having identified a total of 65 papers published between 1983 and mid-2019, including journal articles (27) and working papers/reports (23) addressing case studies in four regions: Africa, Asia, North America, and Latin America. India and Peru are the most represented countries, analyzed in seven and five studies, respectively. Twelve studies address effects on firms' outcomes; the remaining studies use households as the unit of observation. Most of the studies focus on the effects of access to electricity, while six analyze the effects of improvements in the quality of electricity. None of the latter examines Latin American countries.

The *effect size (β)* expresses the reported estimates as relative average changes resulting from the intervention. This is direct, for example, in the case of coefficients resulting from a log-linear specification.[3] The effect size, in relative terms, is usually reported by the Authors, typically with respect to the control mean. For this analysis, the effect size has been calculated (or verified) based on the available information in each study. Since the control mean (or sample sizes) is not available in all cases, the calculation is based on the most appropriate available alternative, typically the sample outcome mean. When the original specification in the studies differs from the log-linear, the effect size's SE is derived from the SE or $t-stat$ reported in the original model. Also, the indicators are adjusted such that all have the same expected direction of impact. Expressing the effect size as relative changes, as well as the mentioned recalculations, facilities comparability across studies with heterogeneous specifications and results (e.g., variables with different metrics or scales; evaluation periods; populations with different variances; and so forth). However, some estimates per study, and 16 complete studies are not included in this analysis due to insufficient information to reconstruct the estimated average effect on an equivalent basis, or because they focus on different outcomes. It is important to emphasize that, for this exercise, no authors were contacted to request their data set or codes.

[3] Other example of effect size measured in relative terms are meta-analyses based on log-log specifications, as for example Espey [1998] for the case of gasoline demand elasticities.

The final data set is composed by 196 estimates from 49 impact evaluations published between 1983 and mid-2019. It includes seven studies with indicators at the aggregate level (e.g., districts, villages) and six studies with firms as the unit of observation (of which 2 use aggregate data). The sample sizes of the studies varies greatly. Studies using indicators at household level as unit of observation pooled over 3 million observations, ranging from 53 to 952,153 observations per study. Studies using indicators at firm level comprise close to 380 thousand observations ranging from 62 to 374,383.[4] The data set contains 23 variables from the studies' findings and their characteristics. Regarding the characteristics, the data set includes type of publication; type of treatment (access, service improvement); type of main independent variable (continuous, discrete); area of study (rural, urban, both); unit of observation; identification strategy; country of study; data structure (panel, cross-section, etc.); evaluation period (measured as the difference between the baseline year and the last follow-up. It is coded as zero for cross-sectional studies); sample size; outcomes evaluated; and outcome differentiation by gender. For each country under study, the data set also merges its corresponding per-capita income calculated from the Penn World tables.

4 Methodology

This section presents a framework in which to analyze the heterogeneity of the compiled estimates, and to understand the extent to which they are informative about the effectiveness of an intervention. Following Stanley and Doucouliagos [2012], assume that the estimate of study i, b_i, is normally distributed around the true treatment effect β with sample variance σ^2, and N the sample size of the study.

$$b \sim (\beta, \sigma^2). \tag{1}$$

The realized value of the intervention can be written as

$$b = \beta + sez, \tag{2}$$

where z is the realization of the normal variate. Dividing by the standard error (se)

$$t_i = \beta/se + z. \tag{3}$$

This equation implies that the observed t-statistics represent the effect size, β/se, adjusted by the sample size plus a realization of a standard normal variate, and no other component should affect the t-statistic. Equation 4 considers, in addition, a constant term

[4] Household studies using aggregate data comprise a sample of around 76 thousand, while firm studies pooled approximately 63 thousand observations

as the simplest case to account for other factors that may be influencing the reported estimates and their corresponding statistical significance.

$$t_i = \beta/se + \alpha + z. \tag{4}$$

If α is significant, that is an indication that the reported estimates systematically deviates from a symmetric distribution. Equation (4), therefore represents the basic test for selective publication by the researcher or by the journal (i.e. p-hacking, specification searching, or publication bias of the journals). In the case of electrification, α is expected to be positive because of the hypothesized direction of the electrification impacts.

Alternatively, Stanley et al. [2007] argue that since the effect of the standard errors is nonlinear,it is better to model the asymmetry by adding a nonlinear term in se

$$t_i = \beta/se_i + \gamma se_i + z. \tag{5}$$

In equation (5), β represents the underlying population effect with γ the quadratic effect of selective publication.

Further, the constant in equation (4) can be decomposed to account for covariates that may explain heterogeneity in the underlying effects, as well as, the reporting bias. On the other hand, notice that in this framework the disturbance term is unlikely to be homoskedastic, i.e., the covariates are directly related to se. Therefore, it is recommended to weight the studies using the inverse-variance weights [Stanley, 2008, Marín-Martínez and Sánchez-Meca, 2010] suggests to implement weighted least squared. Another consideration is that within study estimates can be correlated because of the researcher's estimation methods, data or priors. The intra-study correlation would be greater if the number of estimates per study or researcher were larger. The sample in this study presents 16 studies with two, three, four and six estimates for income, for example. To deal with this problem, the meta-analysis literature recommends treating the estimates from the same study as hierarchical groups or clusters, where typically the regression intercept is modeled as random, such that all estimates are correlated within the group.[5] This model allows us to include additional covariates to characterize the observable heterogeneity. The resulting random mixed effects multilevel model takes the form

$$t_{ij} = \beta/se_{ij} + \gamma se_i + u_i + \varepsilon_{ij} \tag{6}$$

$$t_{ij} = \beta/se_{ij} + \alpha + \theta_k M_{ijk} + \delta_k C_{ijk} + P_i \eta_i + u_i + \varepsilon_{ij} \tag{7}$$

where each study i can have j estimations for a given outcome. As mentioned in the data section, equivalent indicators are grouped into income, labor and educational outcome-

[5] At other levels, such outlet, observation are less frequent

categories. This helps to reduce the relatively small size of the literature. The explanatory variables in Equation (7) are separated into methodological characteristics of the study (M), context-specific characteristics of the case study (C), and publication variables (P). All these variables are divided by se. The vectors of parameters θ, δ_k and η_k capture their respective association with the reported effects.

Specifically, methodological features (M) include a score variable that captures the influence of studies applying methodologies that require weaker identification assumptions. The highly endogenous placement of infrastructure projects poses substantial identification challenges to isolate credible sources of exogenous variation. In the case of electrification, for example, Lee et al. [2019a], highlight the potential importance of methodological approaches to reconcile disparate findings. In general, Brodeur et al. [2019] discuss the sensitivity of the research methods to p-hacking, finding evidence that experimental methods tend to be less prone to such a problem. Typically, different methods rely on different data structures that add to (or enable) the credibility of the proposed identification setting.

Given the scarcity of experimental approaches in the electrification literature, in this paper, the methodological dimension is included as a score that reflects the interaction between the identification method and the data structure. The data structure differentiates between cross-section, pooled cross-section, and panel data. The method-indicator is equal to 1 if the study uses a method different from OLS (i.e. IV, double differences, RDD). Considering the score variable as an interaction of the method and the data-structure, also helps to reduce the co-linearity between both variables, and also deals with the small number of studies using quasi-experimental methods with richer data structures. [6]

M also includes an indicator variable for evaluation periods greater than five years because short exposure periods may make it difficult to observe the materialization of impacts of electrification. For example, income gains or increased schooling related to the productive use of electricity infrastructure would take some time to occur. In fact, some of the larger estimated income effects in the literature are found in studies that address longer evaluation periods such as those of Lipscomb et al. [2013] in Brazil and Chakravorty et al. [2016] in India. To account for potential gender differential effects, M includes a gender categorical variable indicating if the outcome is female specific, male specific or for both (base level). In general, it is expected that M helps us to understand the role of identification problems in explaining the heterogeneity of the observed estimates.

Regarding context-specific characteristics (C), the analysis includes the income level of the country studied, as well as the region and indicators for the specific variables that are grouped into each outcome category. These variables are expected to account for different

[6] The results are consistent when the method variable distinguishes according to the following hierarchy: randomization, regression discontinuity, instrumental variable, double differences, propensity score matching, and before and after (including cross-section and pooled cross-sections).

electrification contexts, specifically, for the fact that electrification does not operate in a bubble, but requires markets and complementary public services to create synergies, and effectively be transformed and used in productive ways. These variables are intended to capture the overall stage of development in the country under analysis, and it is expected that a higher income level is associated with a greater capacity to translate electricity services into positive socioeconomic impacts. I expect C to help explain the variability that may be expected from treating different populations.

The set of publication variables (P) includes an additional indicator for studies published in a peer-reviewed journal, and the year of publication, to further characterize potential selective publication. That is, if publishing in a peer-reviewed journal is associated with our dependent variable, that suggests that publication bias would be driven by pressure on editors or referees. Following Goldfarb [1995], the inclusion of the year of publication helps to explore the presence of time patterns in publication bias, which in this case may be associated with the increasing attention that electrification has received in recent years.

To control for additional differences across studies, regressions also include an indicator for effects calculated for households (vs. firms, applicable for income and labor outcomes); an indicator variable for studies based on an aggregated variable; indicators for the each region in which the studies took place; and specific impact-indicators within each outcome category.

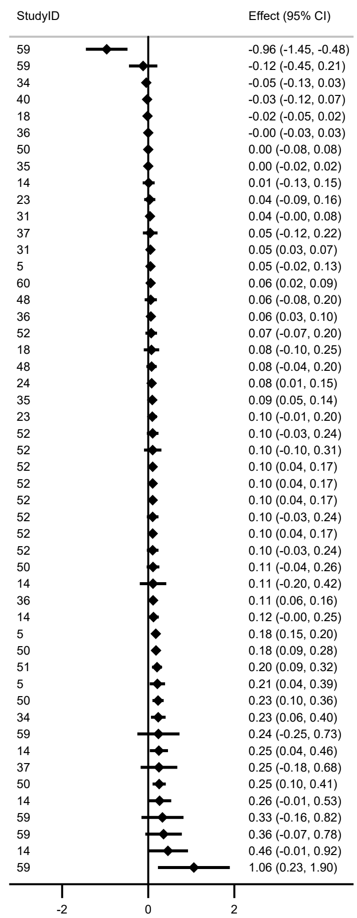
5 Results

5.1 Descriptive review

An overview of the literature and its contrasting results are depicted on the forest plot presented in Figure 2. This figure presents the average estimates and their corresponding 95 percent confidence interval by each study estimate in the sample. These estimates vary widely and are, in many cases, not statistically significant. In the case of income, for example, the minimum effect is around -18 percent, though not significant statistically, and the maximum reported effect is an astonishing 178 percent, statistically significant at 99 percent confidence. For employment, the minimum effect is close to -100 percent and statistically significant at 99 percent. Overall, more than one half and one third of the estimates are significant in the cases of enrollment and years of schooling, respectively. Similarly, around half the estimates indicate non-significant impacts on labor and income outcomes. See forest plot for employment in Figure 2 (also see Figure S1-S4 for other outcomes).

The summary statistics by outcome are presented in Table 1. The average estimated effects (and 95-percent confidence interval (CI)) are in the order of 24 percent (CI=0.15-

Figure 2: Forest Plot of Employment



0.35), 17 percent (CI=0.02-0.34), eight percent (CI=0.04-0.13) and ten percent (CI=0.04-.16) for income, labor, enrollment and years of schooling, respectively. Given the substantial variation observed within and across studies, however, the last column of Table 1 presents the standardized effects (i.e., the effects divided by their standard deviation), which is similar to the Cohen-d typically reported in meta-analysis. Consistent with the observed, pronounced variance in the literature estimates, the results show smaller actual effects according to standard criteria [Cohen, 1988]. That is, for all outcomes, the estimated effects are less than a 0.20 standard deviation.

In the appendixes, Table A1 presents the breakdown by gender and evaluation period, consistently showing high heterogeneity in the estimates. Further, the break- down shows larger effects for income and labor when the outcome is female-specific and, contrary to expectations, smaller effects for studies with larger evaluation periods.

A relevant question is whether the observed heterogeneity reflects variation in true effects rather than sampling errors. The I^2 index provides an approximation of the proportion of the variation in true effects that is contained in the variation of the observed estimates [Higgins et al., 2003, Borenstein et al., 2017]. According to this measure, 80 percent or more of the observed heterogeneity could be attributed to an underlying true effect dispersion. This suggests that the size of the effects would actually vary among populations, while the sampling error would play a minor role.

Table 1: Descriptive Stats

	Obs.	Mean effect	Min effect	Median effect	Max effect	SD effect	Weighted Mean	Weighted β/SD_β
Income	57	.248	-.182	.161	1.78	.37	.235	.033
Time Worked	23	.327	-.454	.042	5.53	1.16	.28	.022
Employment	50	.114	-.965	.103	1.06	.232	.127	.019
Schooling	28	.103	-.347	.156	.314	.155	.134	.048
Enrolment	26	.084	-.043	.054	.51	.122	.061	.031
Study time	12	.411	-.036	.273	1.5	.444	.527	.071

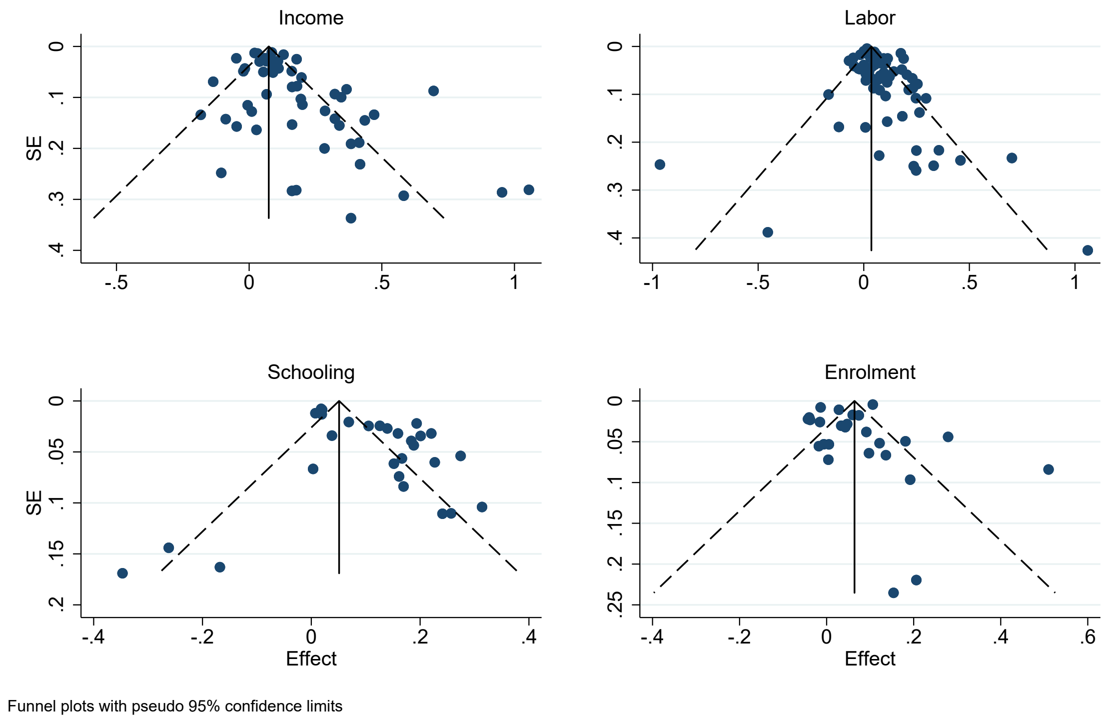
Note: The mean, min., median, max., and standard deviation (SD) are reported for unweighted estimated effects. In the case of the β/SD_β, there is one observation less for income and labor outcomes, and two fewer observations for educational outcomes due to non identified sample size in the corresponding regressions of three studies. The β/SD_β is calculated based on the weighted estimates. See breakdown by gender and evaluation period in Appendix A1.

5.2 Graphical analysis of Publication Bias

As a visual inspection whether the contrasting results in the literature may be, to some degree, explained by publication bias, Figure 3 presents the distribution of the estimates using the Funnel plot, with the mean effect in the x-axis and a measure of the study precision in the y-axis (inverse standard error in this case). The Funnel approach assumes that higher-powered studies should have estimates closer to the true effect, and should be located higher in the figure, while lower-powered studies should have more sparse estimates at the bottom of the figure. In the absence of publication bias, the distribution of the pairs (effect, precision) should be symmetrical. On the contrary, in the presence of publication bias, studies with small sizes may tend to present larger effects, skewing the distribution of the Funnel plot. The plots include pseudo-95percent confidence interval lines, which are drawn around the summary fixed-effect estimate of the intervention or treatment effect. According to Figure 3, the asymmetric distribution of the effects is apparent in the sample under analysis. The pairs, effect-precision, are right-skewed, indicating possible publication bias in income, employment and, less clearly, enrollment and schooling (for which the samples are smaller).

Since the Funnel plot is a test of symmetry, the fill-and-trim method can be used to estimate the number of studies that are likely missing, in order to balance the distribution [Duval and Tweedie, 2000a]. According to Duval and Tweedie [2000b], under this method, studies could be missing for different reasons; including publication bias and poor study design, among others. This analysis suggests that 17, 22, seven, and one studies or estimates would be missing from the Funnel plot for income, labor schooling, and enrollment, respectively. The inclusion of these studies/estimates would reduce the overall effect for all outcomes by statistically significant levels, except for enrollment.[7]

Figure 3: Funnel Plots



5.3 Meta-Regressions

Test of Selective Publication: consistent with previous examination, the publication bias parameter (α in equation 4) is statistically significant at the one percent level for all outcomes (see Table 2-Panel A). Along with the results from the fill-and-trim method, these results suggest that smaller and non-significant effects tend to be reported less often in the literature. That is, starting from the general conceptual framework, lay-out in section 2, these results could be interpreted as the presence of confirmation bias.

Underlying Effects of Electrification: to analyze the extent to which publication bias may be masking the underlying population effects in this sample, Table 2-Panel B presents

[7]Under the random effect model, the pooled estimates go from 0.12 to 0.07, from 0.07 to 0.04, from 0.12 to 0.07, and from 0.06 to 0.05 for income, labor, schooling and enrollment, respectively. In the last case, the 95 percent confidence intervals overlap.

the estimation results of equation 6, in which the underlying effect is captured by coefficient β and, as conceptually expected, is positive and significant at the one percent level for all outcomes. Notice, these estimates are consistently smaller than the average (and median) of this sample (see Table 1). The estimated underlying effect is 6.9 percent for income, 5.3 percent for employment and 2.6 percent for enrollment (versus the weighed averages of 24 percent, 13 percent and 6 percent, respectively).

Table 2: Meta-Regressions

	(1) Income	(2) Time paid work	(3) Employment	(4) Schooling	(5) Enrolment
Panel A: Selective Publication					
Precision	0.042***	0.009	0.027**	0.008	0.017**
	(0.012)	(0.007)	(0.011)	(0.012)	(0.008)
Pub. Bias	1.445***	1.657***	1.467***	2.414***	3.323**
	(0.346)	(0.386)	(0.376)	(0.653)	(1.593)
Panel B: Underlying Effects					
Underlying Effect	0.069***	-0.009	0.053***	0.034***	0.026***
	(0.010)	(0.007)	(0.008)	(0.011)	(0.009)
SE	1.833*	2.909***	7.456***	15.502**	18.869*
	(0.989)	(1.097)	(2.265)	(7.570)	(10.991)
N	57	23	50	28	26

Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Augmented Meta-Regressions: to complement previous results, Table 3 shows the estimations from equation (7), allowing us to analyze which study characteristics may account for the variation in the reported effect size. Notice that in these specifications, publication bias should be read as the joint statistical significance of three variables (i.e., if published in a journal, year of publication and the constant term). The corresponding joint test suggests the presence of publication bias in all of the outcome categories at the five-percent level. In particular, in all cases the constant terms is strongly associated with the standardized effects, while being published in a journal is positively correlated with the effects in all cases, but strongly only in the case of labor and educational outcomes. That is, confirmatory bias seems to be present both among authors and journals, respectively. As per year of publication, there is no clear direction in the time-pattern of publication bias.

Regarding the set of methodological characteristics of the studies, the results also indicate that more rigorous empirical approaches and rich data-structures are systematically associated with more-conservative estimates of the impact of electrification. The potential role of differences in empirical methods has also been stressed by Lee et al. [2019a], highlighting the challenges in cleanly identifying causal links in electrification programs, and noticing that more recent studies, using more rigorous approaches, tend to find smaller impacts [e.g. Burlig and Preonas [2016], Lee et al. [2019b]].

Regarding the context of the studies, the results indicate that they help to explain the heterogeneity in the literature's findings, though without a clear direction and, in some cases, only at a weak statistical significance level. Long-run evaluation periods are, for example, only strongly correlated with the educational outcome category. On the other hand, contrary to expectations, higher income per-capita (of the country in which the study took place) is positively associated with educational impacts, but negatively associated with the labor and income dimensions.

The results also suggest that electrification effects tend to be heterogeneous across gender. Recall that for these results the base level is the outcomes that does not distinguish gender. That is, male outcomes seem not to have differentiated effects from the average estimated effect in any outcome category. In contrast, the estimations suggest that female-specific outcome are weakly positively correlated with greater reported effects, particularly in the labor and educational categories for women and girls, respectively. This result seems consistent with a conceptual framework in which women and girls tend to be the primary beneficiary of the improved dwelling infrastructure, potentially boosting their local productive opportunities.

Overall, the analysis suggests that the electrification effects (in the evaluated outcomes) are smaller than those which could have been anticipated by the weighted averages of the pooled studies, and that they are truly heterogeneous. These smaller effects could be regarded as more plausible since electricity represents only one dimension of development. The productive opportunities emerging from electricity infrastructure can only do so much, as they also depend on factors such as access to credit; market depth, maturity and access; and complementary infrastructure, among others. In fact, the noted relevance of methodological set-ups seems to back up this plausibility claim, as more stringent methods help to obtain cleaner–and less sizeable–estimated effects of electrification alone. However, it is important to emphasize the difficulty in disentangling the role of context-specific variables. On the one hand, the number of studies is still relatively small to identify a variety of factors and their interactions (e.g. access to electricity and access to road, financing, etc.). On the other hand, there is room for improvement in terms of capturing relevant contextual factors particular to the intervention (e.g. quality of electricity provided, cost of the services, etc.).

Table 3: Augmented Meta-Regressions

	(1) Income	(2) Labor	(3) Education
t stat			
Precision	-0.075	-0.001	-0.222***
	(0.304)	(0.256)	(0.084)
Methodology X Data Struc.	-0.057**	-0.023***	-0.027***
	(0.026)	(0.009)	(0.006)
Long-run evaluation	0.041	0.013	0.092***
	(0.045)	(0.019)	(0.015)
GDP per capita	-0.031*	-0.042**	0.063***
	(0.017)	(0.019)	(0.011)
Female outcome	0.025	0.029*	0.027*
	(0.048)	(0.016)	(0.014)
Male outcome	0.016	-0.021	0.015
	(0.042)	(0.016)	(0.014)
Published	0.011	0.031*	0.063***
	(0.019)	(0.017)	(0.017)
Year of publication	0.003	0.003**	-0.010***
	(0.004)	(0.002)	(0.002)
Constant	1.967***	1.554***	1.401***
	(0.317)	(0.236)	(0.273)
Observations	57	73	66

Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Note: All regressions control for region of the country where the case-study took place; indicator variable for household outcomes; indicator variable for studies with aggregated observation-units; and specific impact-indicators within each outcome category. See a detailed breakdown by specific impact-indicator in Appendix A2.

6 Conclusion

This paper synthesizes econometric evaluations of electrification interventions using a new and comprehensive sample of program estimates over the period 1984 to June-2019. The analysis points to a number of important lessons for energy policy-making, and infrastructure program evaluation.

First, the literature's findings regarding the socioeconomic effects of electrification are substantially heterogeneous and still inconclusive regarding both their statistical significance and materiality. Accounting for such variability in the estimates, the average effect size (in the analyzed variables) is less than a 0.2 standard deviation, typically considered small in the meta-analysis literature.

Second, the actual reporting of estimates seems to be influenced by selective publication bias. After accounting for this source of bias, the underlying estimated effects are still positive and statistically significant, but substantially smaller than the simple or weighted averages. It is worth-mentioning that, regardless the wide variability in the estimates, the direction of the bias is consistent with what would be expected from the conceptual standpoint, suggesting that reported estimates may be trying to confirm a working hypothesis of positive electrification effects.

Third, the variability in the literature's estimates can be explained, to an extent, by methodological approaches and context-specific characteristics of the evaluations. Approaches imposing weaker identification assumptions seems to be associated with the significance and sign of the estimates. Nonetheless, this examination also suggests that there is substantial heterogeneity in underlying effects. Effects seems to be heterogeneous across gender and context variables of the case study, such as country income per-capita and length of the evaluation period.

The present synthesis reviews a number of studies consistent with previous meta- analyses. However, impact evaluation studies on electrification are relatively scarce in comparison to other topics (e.g., gasoline elasticity of demand, labor market program evaluations). The relatively small sample size should be taken into account in carefully weighting the results presented here. This scarcity reflects the challenges of performing impact evaluations of electricity systems, where there tend not to be markets, or it is difficult to sort treatment on a comparable basis. The sample size, along with reporting limitations of the reviewed studies, also limit the capacity to explore interactions between the factors that characterize the interventions analyzed.

As such, it is valuable to note some dimensions along which the literature can improve. First, studies focus mostly on impacts, and to a lesser extent on channels of how electrification actually delivers. This is to be expected given the difficulty of evaluating the effects of electrification, however, it also represent a limitation to understand to role of electricity in the development mechanism. Second, the reporting of information across studies is also quite heterogeneous, thus limiting the comparability of estimates. Third, the cost side of the interventions is mostly absent in the literature, precluding assessments of their costs-benefits. Overall, these limitations represent a knowledge gap that directly reduces the capacity for designing better public policies, as they limit our understanding of the development process that is enabled as a result of access to modern energy sources, as well as their cost-weighted effects. In a policy environment in which electrification has emerged as a key development factor, a take-away is that greater support could be posed to strategically position the practice of rigorous analyses of its effectiveness and costs-benefits.

References

Manuel Barron and Maximo Torero. Household electrification and indoor air pollution. *Journal of Environmental Economics and Management*, 86:81–92, 2017.

Michael Borenstein, Julian PT Higgins, Larry V Hedges, and Hannah R Rothstein. Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. *Research synthesis methods*, 8(1):5–18, 2017.

A. Brodeur, N. Cook, and A. Heyes. Methods matter: P-hacking and causal inference in economics and finance. *Working paper*, 2019.

Fiona Burlig and Louis Preonas. Out of the darkness and into the light? development effects of rural electrification. *Energy Institute at Haas WP*, 268, 2016.

Ujjayant Chakravorty, Kyle Emerick, and Majah-Leah Ravago. Lighting up the last mile: The benefits and costs of extending electricity to the rural poor. *Resources for the Future Discussion Paper*, pages 16–22, 2016.

J Cohen. Statistical statistical power analysis for the behavioral sciences, 1988.

Esther Duflo. Richard t. ely lecture: The economist as plumber. *American Economic Review*, 107(5):1–26, 2017.

Sue Duval and Richard Tweedie. A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, 95 (449):89–98, 2000a.

Sue Duval and Richard Tweedie. Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2):455–463, 2000b.

Molly Espey. Gasoline demand revisited: an international meta-analysis of elasticities. *Energy Economics*, 20(3):273–295, 1998. ISSN 01409883. doi: 10.1016/S0140-9883(97)00013-3.

Karen Fisher-Vanden, Erin T Mansur, and Qiong Juliana Wang. Electricity shortages and firm productivity: evidence from china's industrial firms. *Journal of Development Economics*, 114:172–188, 2015.

Paul J Gertler, Sebastian Martinez, Patrick Premand, Laura B Rawlings, and Christel MJ Vermeersch. *Impact evaluation in practice.* The World Bank, 2016.

Robert S Goldfarb. The economist-as-audience needs a methodology of plausible inference. *Journal of Economic Methodology*, 2(2):201–222, 1995.

Louise Grogan. Time use impacts of rural electrification: Longitudinal evidence from guatemala. *Journal of Development Economics*, 135:304–317, 2018.

Julian PT Higgins, Simon G Thompson, Jonathan J Deeks, and Douglas G Altman. Measuring inconsistency in meta-analyses. *Bmj*, 327(7414):557–560, 2003.

IEA. *World Energy Outlook*. "International Energy Agency (IEA)", 2018.

K. Lee, Edward Miguel, and C. Wolfram. Does electrification supercharge economic development? *Working paper*, 2019a.

Kenneth Lee, Edward Miguel, and Catherine Wolfram. Experimental evidence on the economics of rural electrification. *Journal of Political Economy*, 0(ja):null, 2019b. doi: 10.1086/705417. URL `https://doi.org/10.1086/705417`.

Molly Lipscomb, A Mushfiq Mobarak, and Tania Barham. Development effects of electrification: Evidence from the topographic placement of hydropower plants in brazil. *American Economic Journal: Applied Economics*, 5(2):200–231, 2013.

Fulgencio Marín-Martínez and Julio Sánchez-Meca. Weighting by inverse variance or by sample size in random-effects meta-analysis. *Educational and Psychological Measurement*, 70(1):56–73, 2010.

Juan Pablo Rud. Electricity provision and industrial development: Evidence from india. *Journal of development Economics*, 97(2):352–367, 2012.

TD Stanley, Chris Doucouliagos, et al. Identifying and correcting publication selection bias in the efficiency-wage literature: Heckman meta-regression. *Economics Series*, 11:2007, 2007.

Tom D Stanley. Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection. *Oxford Bulletin of Economics and statistics*, 70(1): 103–127, 2008.

Tom D Stanley and Hristos Doucouliagos. *Meta-regression analysis in economics and business*, volume 5. Routledge, 2012.

Maximo Torero. The impact of rural electrification: Challenges and ways forward. *Revue d'économie du développement*, 23(HS):49–75, 2015.

7 Supplementary Material

Figure S1: Forest Plot of Time Paid Work

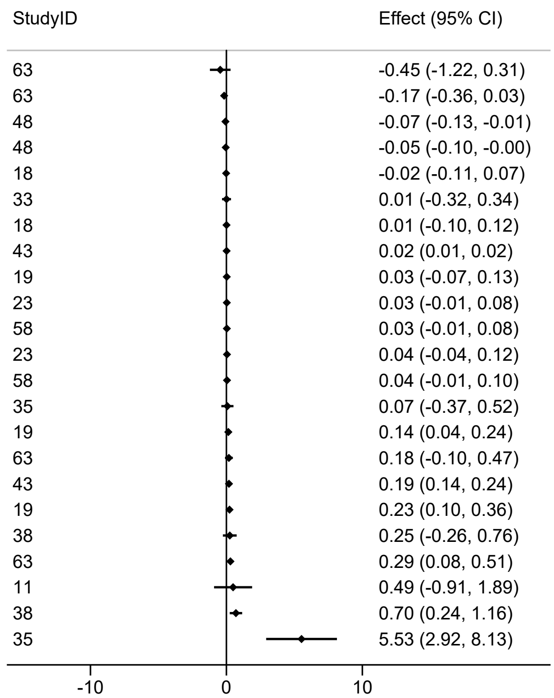


Figure S2: Forest Plot of Income

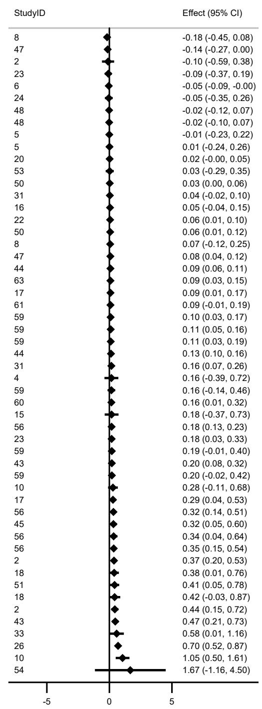


Figure S3: Forest Plot of Schooling

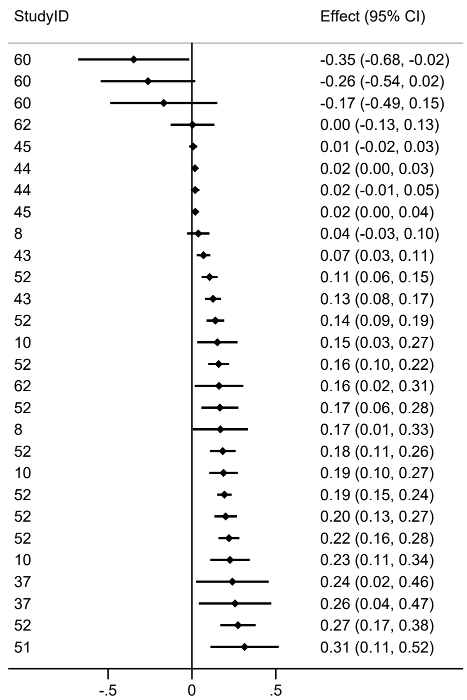


Figure S4: Forest Plot of Enrolment

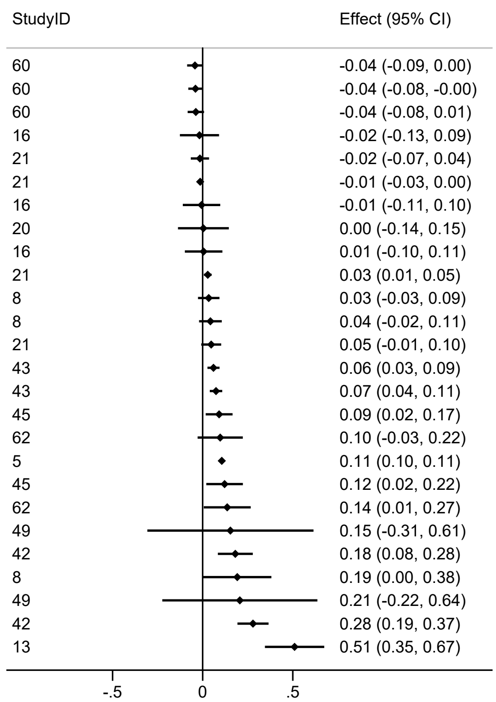


Table A1: Descriptive Statistics by Outcome

	Female				Male				Short-run				Long-run			
	Mean	P5	P95	Obs	Mean	P5	P95	Obs	Mean	P5	P95	Obs	Mean	P5	P95	Obs
Average																
Income	0.23	0.23	0.23	6	0.23	0.23	0.23	5	0.23	0.23	0.23	29	0.23	0.23	0.23	28
Time paid work	0.28	0.28	0.28	10	0.28	0.28	0.28	8	0.28	0.28	0.28	13	0.28	0.28	0.28	10
Employment	0.13	0.13	0.13	15	0.13	0.13	0.13	13	0.13	0.13	0.13	19	0.13	0.13	0.13	31
Schooling	0.13	0.13	0.13	10	0.13	0.13	0.13	10	0.13	0.13	0.13	11	0.13	0.13	0.13	17
Enrolment	0.06	0.06	0.06	8	0.06	0.06	0.06	8	0.06	0.06	0.06	16	0.06	0.06	0.06	10
Study time	0.53	0.53	0.53	1	0.53	0.53	0.53	1	0.53	0.53	0.53	12	.	.	.	0
Cohen d																
Income	0.03	0.03	0.03	6	0.03	0.03	0.03	5	0.03	0.03	0.03	29	0.03	0.03	0.03	28
Time paid work	0.02	0.02	0.02	10	0.02	0.02	0.02	8	0.02	0.02	0.02	13	0.02	0.02	0.02	10
Employment	0.02	0.02	0.02	15	0.02	0.02	0.02	13	0.02	0.02	0.02	19	0.02	0.02	0.02	31
Schooling	0.05	0.05	0.05	10	0.05	0.05	0.05	10	0.05	0.05	0.05	11	0.05	0.05	0.05	17
Enrolment	0.03	0.03	0.03	8	0.03	0.03	0.03	8	0.03	0.03	0.03	16	0.03	0.03	0.03	10
Study time	0.07	0.07	0.07	1	0.07	0.07	0.07	1	0.07	0.07	0.07	12	.	.	.	0

Note: Statistics calculated from weighted means.

Table A2: Augmented Meta Regressions by Outcome

	(1) Income	(2) Time Workerd	(3) Employment	(4) Schooling	(5) Enrolment
t stat					
Precision	-0.075	-0.351	-0.321	0.031	0.921***
	(0.304)	(0.468)	(0.377)	(0.244)	(0.251)
Methodology X Data Struc.	-0.057**	0.070*	-0.011	-0.036***	0.004
	(0.026)	(0.037)	(0.016)	(0.008)	(0.008)
Long-run evaluation	0.041	-0.230***	0.024	0.097**	-0.049
	(0.045)	(0.085)	(0.022)	(0.047)	(0.037)
GDP per capita	-0.031*	0.007	0.044	0.056***	-0.069**
	(0.017)	(0.055)	(0.048)	(0.021)	(0.033)
Female outcome	0.025	0.128***	0.016	-0.023	0.004
	(0.048)	(0.027)	(0.022)	(0.017)	(0.010)
Male outcome	0.016	0.125***	-0.048***	-0.034**	-0.011
	(0.042)	(0.035)	(0.016)	(0.017)	(0.010)
Published	0.011	0.067*	0.033	0.109**	0.072
	(0.019)	(0.037)	(0.024)	(0.045)	(0.047)
Year of publication	0.003	-0.001	0.001	-0.018*	-0.008***
	(0.004)	(0.008)	(0.002)	(0.010)	(0.003)
Constant	1.967***	1.904***	0.948**	2.112***	0.704*
	(0.317)	(0.598)	(0.378)	(0.400)	(0.377)
Observations	57	23	50	28	26

Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$