

## Graphical models and the challenge of evidence-based practice in development and sustainability

Ryan S.D. Calder<sup>a\*</sup>, Andrea Alatorre<sup>b</sup>, Rebecca S. Marx<sup>b</sup>, Varun Mallampalli<sup>a</sup>, Sara A. Mason<sup>c</sup>, Lydia P. Olander<sup>c</sup>, Marc Jeuland<sup>d</sup>, Mark E. Borsuk<sup>a</sup>

<sup>a</sup> Department of Civil and Environmental Engineering, Pratt School of Engineering, Duke University, Durham NC 27708

<sup>b</sup> Nicholas School of the Environment, Duke University, Durham NC 27708

<sup>c</sup> Nicholas Institute for Environmental Policy Solutions, Duke University, Durham NC 27708

<sup>d</sup> Sanford School of Public Policy, Duke University, Durham NC 27708

\* Corresponding author: 1116 Hudson Hall, Box 90287; email: [ryan.calder@duke.edu](mailto:ryan.calder@duke.edu); phone: (919) 660-6883; fax: (919) 660-5219

### Abstract

Governments and social benefit organizations are expected to consider evidence in decision-making. In development and sustainability, evidence spans disciplines and methodological traditions and is often inconclusive. Graphical models are widely promoted to organize interdisciplinary evidence and improve decision-making by considering mediating variables. However, the reproducibility, objectivity and benefits for decision-making of graphical models have not been studied. We evaluate these considerations in the setting of energy services in the developing world, a contemporary development and sustainability imperative. We develop a database of relevant causal relations (331 concepts, 1,355 relationships) asserted in the literature (561 peer-reviewed articles). We demonstrate that high-level relationships of interest to practitioners feature less consistent evidence than the causal relationships that underpin them, supporting increased use of problem decomposition through graphical modelling approaches. However, adding such detail increases complexity exponentially, introducing a hazard of overparameterization if evidence is not available to match the level of mechanistic detail.

### Keywords

results chain; Bayesian network; logic model; evidence assessment

### Version notes

Manuscript accepted for publication in *Environmental Modelling & Software* (last revised by authors 2020-03-31). The final published article may differ from this version.

Article DOI: [10.1016/j.envsoft.2020.104734](https://doi.org/10.1016/j.envsoft.2020.104734)

Available online: 2020-05-06

### Copyright notice

© 2020. This manuscript version is made available under the CC-BY-NC-ND 4.0 license.

Details: <http://creativecommons.org/licenses/by-nc-nd/4.0/>



## 1. Introduction

‘Evidence-based practice’ has become a mantra of international development, social policy, environmental management and many other fields.<sup>1-4</sup> In sustainability and development, increased awareness of policy failures has heightened expectations of accountability on the part of governments and social benefit organisations.<sup>5</sup> Furthermore, expectations of ‘evidence’ for interventions and public policy-making have evolved similarly to those in the clinical, medical and laboratory sciences. The 20<sup>th</sup> century saw the evolution of medicine in particular from a collection of physician-centered traditions and semi-scientific beliefs<sup>2,6</sup> to a positivist ‘orthopraxy’ undergirded by ‘causal chains’ between intervention and outcome.<sup>7</sup> This transition was facilitated by experimental methods (in particular the randomized controlled trial) and evidence synthesis centered around the systematic review.<sup>2</sup>

These expectations of greater methodological rigor have been adopted by diverse fields spanning the natural, applied and social sciences, even where randomization and experimentation are impossible. Frameworks for systematic reviews of non-clinical evidence such as Campbell (social interventions) and 3ie (international development interventions) strive to represent the weight of evidence on ‘what works’ from accumulated studies, frequently across diverse social and environmental contexts.<sup>8</sup> Systematic reviews in development and sustainability however face a number of challenges relating to inter-study variability and causal complexity less commonly faced by the laboratory sciences. Notably, interventions are rarely identical in all respects and impact unique socio-environmental processes lacking an inherent beginning or end. These factors often preclude systematic reviews of the social sciences from conclusively identifying common modes of action, causal structures or relevant supporting or impeding factors.<sup>2,9</sup>

Mallett et al.<sup>10</sup> describe how systematic reviews to inform international development tend to favor studies of high-level associations and omit research on the underlying socio-environmental factors that mediate those associations. Other authors have highlighted how the frequent inconclusiveness of systematic reviews often makes them more valuable in the identification of research gaps than for the purposes of informing public policy.<sup>10,11</sup> In settings where causal structures span academic disciplines and communities of practice, evidence synthesis and identification of causal pathways are further hampered by a fragmented literature<sup>12</sup> and by modelling tools that do not adequately capture interactions between social and physical

systems.<sup>13,14</sup> Traditional methods for evidence synthesis have therefore had a mixed record with respect to averting policy failures and unintended consequences. For example, Hartvigsson et al.<sup>15</sup> describe how the inadequate conceptualization of unique socio-environmental systems has contributed to inconsistent outcomes of rural electrification.

Graphical modelling has the potential to improve evidence synthesis by depicting the web of relations uncovered in a systematic review as an articulated series of causal claims linking interventions to outcomes.<sup>15,16</sup> In the context of international development and sustainability, for example, Tallis et al.<sup>17</sup> have proposed a ‘results chains’ framework which links graphical modelling methods guidance on evidence evaluation. This framework aims to improve decision-making regarding complex systems by characterizing associations between intervention and outcome in terms of interpretable ‘mechanisms’ and by avoiding disciplinary biases toward certain types of evidence. In the setting of public health, the United States Centers for Disease Control and Prevention<sup>18</sup> propose to structure action, outcome and mediating variables within ‘logic models’ that they believe ‘increase the likelihood that program efforts will be successful’.

However, the optimal level of detail and the relevant temporal and spatial scales for such models are rarely interrogated directly. In particular, it is unknown to what extent increased complexity in model structures improves mechanistic understanding of the interactions in socio-environmental systems or reduces predictive uncertainty. In many fields, ‘mechanisms’ are as much a function of the research questions posed and the disciplinary orientation of researchers as they are a function of underlying system dynamics.<sup>19,20</sup> For example, it is understood that progress on the ‘wicked problems’ facing development, environment, and health<sup>21</sup> has long been complicated by uncertainty and disagreement regarding the appropriate scale of causal mechanisms.<sup>13,14,22</sup> As previously disparate disciplinary evidence is increasingly structured into ‘unified’ graphical models in an attempt to tackle these problems, differences in operational definitions and scales of interest may result in overall structures that are overparameterized relative to their underlying supporting evidence.

Here, we report on a structured literature review that catalogues relationships between energy access interventions in the developing world and diverse development, environment and health outcomes, and the role of socioenvironmental variables in mediating these relationships. After characterizing the distribution of factors and relationships addressed by this literature, we

assemble them into an integrated graphical network, which we compare with expert opinion. By evaluating this network at various levels of detail, we can evaluate the degree to which more mechanistic detail helps resolve conflicting evidence versus leading to model intractability and overfitting. We conclude by reviewing approaches to simplifying complex networks into tractable modelling tools and identifying potential causal mechanisms that lack empirical research.

## 2. Methods

Here we present an overview of the methods used in this work. We include more detailed methods including summary tables and search syntaxes in the supplemental information (SI) as outlined below.

### 2.1. Literature review and network development

We conducted a review of the peer-reviewed literature using Web of Science. We searched for all review articles mentioning energy access and at least one of the following categories of outcomes: environment, development or health. The search was restricted to articles identifying a country in the developing world or the developing world in general. This process returned 1,983 unique review articles, of which 561 were found to be relevant to energy access, development, environment or health concepts in the developing world. Expanded methods for the Web of Science search and selection process are provided in SI **Sections 1.1–1.2**. The search syntax used for all Web of Science searches is included in SI **Appendix A**. An article selection tree is included as SI **Figure SM1**.

Articles were saved in Mendeley<sup>23</sup>. Metadata of relevant articles was extended to include the cause-and-effect relationships asserted or studied by the authors of individual review articles. These relationships were coded using the authors' original terminology and were based on associations described in abstracts (233 articles) or full texts (332 articles) depending on our ability to determine from the abstract the variables studied by each paper. This process is described in greater detail in the supplemental methods (SI **Sections 1.3–1.4**).

Article metadata was then extracted from Mendeley and organized in an R workspace<sup>24</sup> with the RSQLite package.<sup>25</sup> Networks of relationships were generated and queried in R using iGraph.<sup>26</sup> From these articles, we extracted 331 unique concepts connected through 1,355 relationships. A

relationship between two concepts  $A \rightarrow B$  is generated when at least one article claims an association between **A** and **B** and the plurality of such articles claims that **A** has a positive or negative effect on **B**.

## 2.2. Evidence characterization

Of the 1,355 relationships identified in this study, 29 of the most highly studied were assessed in more detail to determine type, breadth and consistency of the underlying evidence base.

Evidence characterizations for direct relationships were completed in approximately decreasing order of the number of times they were identified in the literature review while ensuring representation of diverse topical areas and levels of complexity of underlying networks. We completed as many evidence characterizations as time and resources for this work allowed.

The rubric of Tallis et al. <sup>17</sup> was modified to support a non-site-specific review and implemented for each of the 29 relationships that underwent an evidence characterization. For each relationship, we counted the number of review articles in our database that invoke it and then counted the number of underlying citations (across all review articles in our database) providing each of various types of evidence: correlational data, experimental data, longitudinal data, modelling results or theoretical inference. We counted the number of review articles presenting ‘conflicting’ evidence corresponding to occurrences of dissent from the prevailing direction of effect (including null or inconclusive results of an analysis and considering all time and spatial scales). Detailed methods for the evidence characterization are presented in SI **Sections 1.5–1.6**.

## 2.3. Expert elicitation

We sought to evaluate the general structure of the network we developed from the literature with respect to expert opinion. To do this, we recruited participants from the 2018 Sustainable Energy Transitions Initiative (SETI) conference at Duke University. We recruited participants via email and achieved a 47% enrolment rate (30 enrollees out of 64 conference participants). We excluded current Duke University personnel participating at the SETI conference from the population of participants contacted. Participant demographic information is included in SI **Section 2.1**.

Expert elicitation interviews started with a brief, standardized explanation of conceptual modelling, an explanation of participants’ right to opt out and an explanation of what types of

anonymized data we were likely to publish. Participants then signed consent forms. We then asked them to participate in a ‘guided’ and an ‘unguided’ exercise. In the ‘guided’ exercise, participants selected from a standardized pre-defined list of concepts and drew causal connections between them. In the ‘unguided’ exercise, participants drew similar diagrams related to their own area of research or practice. The final diagrams were photographed and coded by a research assistant (SP). All interviews were recorded. Detailed interview methodology is provided in SI **Section 2.2**

Research ethics approval was granted by the Duke University Institutional Review Board (protocol no. 2018-0497). The Duke IRB reviewed all materials that were used for recruitment and administration of the expert elicitation study. All recruitment and interview materials (scripts, consent forms, etc.) are included in SI **Appendix B**.

### 3. Results

#### 3.1. Literature review

**Figure 1** provides an overview of the distribution of the most common (A) concepts and (B) relationships identified in the literature review. Individual concepts are mentioned in a median of eight review articles, while pairwise relationships are identified in a median of one article (i.e., at least half of all identified relationships are identified only once across 561 articles). About a third of the articles reviewed only address the relationship between two factors, and very few consider causal pathways longer than two relationships (**Figure 2**).

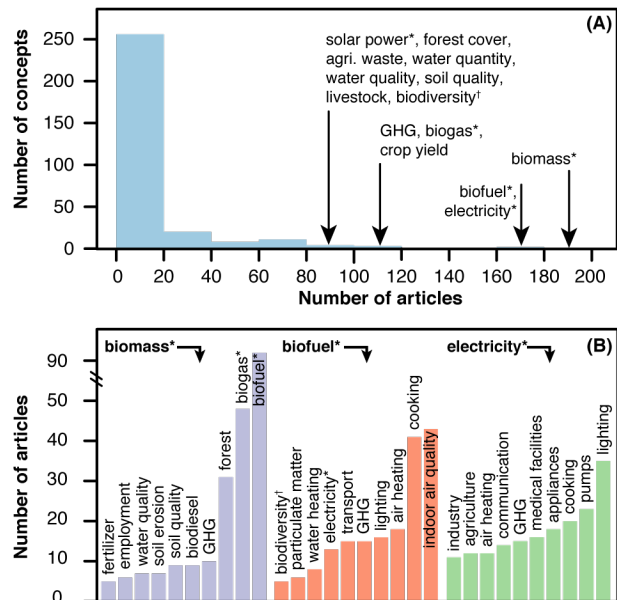


Figure 1: Overview of review articles showing (A) frequency of concepts according to number of articles that invoke that concept, and (B) number of articles describing relationships with the three most commonly invoked concepts. \*Concept included in list of ‘energy’ keywords. † Concept included in list of ‘environment’ outcome keywords.

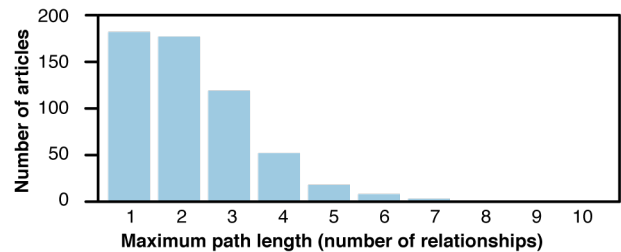


Figure 2: Scope of individual review articles as indicated by the frequency of articles describing paths of various maximum lengths, where path length is defined as the number of relationships in series (for example, A → B → C is length 2).

Certain areas of inquiry are highly represented, with many high-occurrence concepts and relationships were not specifically targeted in the initial search keywords. Roughly half address transitions within or away from biomass or biofuel-based energies (48% of articles coded with *biomass*, *biofuel* or *biogas*). *Electricity*, *crop yield* and *greenhouse gases* were all identified as concepts in >20% of reviewed articles. Possible uses of energy (e.g., *pumps*, *heating*), socio-development outcomes (e.g., *employment*), and environmental and environmental health impacts (e.g., *biodiversity*, *particulate matter*) are relatively highly represented.

Abstracts of reviewed articles were searched for names of countries (using formal names and common variants). The most frequently mentioned countries were China (149 articles) and India (101 articles). A further 15 countries were mentioned in >10 articles. SI **Figure S1** plots the occurrence of all countries appearing in >5 article abstracts and reveals coverage across Africa and Asia. In South America, only Brazil (44 abstracts) and Peru (7 abstracts) appear >5 times. Certain countries appeared in numerous abstracts in the context of international partnerships with industrialized countries (e.g., USA).

### 3.2. Expert corroboration

Expert elicitation interviews produced a similar distribution of concepts and relationships as the literature search. Among all relationships drawn by experts, 70% were reflected in the literature-based network either exactly or via an intermediate concept (i.e., expert drew  $A \rightarrow C$  and literature network contains  $A \rightarrow B \rightarrow C$ ). There is a positive and significant ( $p < 0.001$ ) correlation between the percentage of experts and the percentage of articles invoking given relationships ( $R^2 = 0.05$ ) and concepts ( $R^2 = 0.22$ ). **Figures S2** and **S3** in the supplemental information (SI) plot concepts and relationships, respectively, according to expert invocation vs. literature prevalence.

In comparison to the literature, expert attention focused more on access to lighting and education and costs in relation to income, and less on provision of (or transitions away from) biofuels. In the expert elicitation exercise, 11% of relationships were identified by more than one expert but not identified by the literature review (including by way of an intermediate concept). Most of these reflect the scope of the review. For instance, many experts proposed relationships between

individual diseases and mortality. Our literature search did not target research on specific diseases, so we identified little evidence on these effects (e.g., increased risk of mortality).

Conversely, experts identified a number of pathways that do fall within the scope of our review, but that were not identified from the literature. We plotted a network of all relationships identified by experts that did not appear in the review to identify pathways that may be under-researched or that may reflect gaps in the literature; SI **Figure S4** plots this network of expert-only relationships (excluding relationships identified by only one expert and relationships that exist in our review via an intermediate node, i.e.,  $A \rightarrow C$  is excluded if  $A \rightarrow B \rightarrow C$  exists in our review). Notable pathways include *land rights*  $\rightarrow$  *crop yield*, *food preservation*  $\rightarrow$  *income* and impacts of medical facilities on prevalence of a range of diseases and mortality (e.g., *medical facilities*  $\rightarrow$  *eye disease*). Several other pathways appearing in SI **Figure S3** were coded in the literature review in the opposite direction, for instance, the influence of stove type on fossil fuel use (rather than vice versa).

### 3.3. Graphical network

Most concepts are part of multiple causal pathways and have been studied as initiating events, intermediate effects, and final outcomes. For example, electricity supply has been extensively studied as a predictor of outcomes such as income and intermediate effects such as crop yield, while electricity demand has been studied as an outcome in relation to predictors such as rising incomes. Meanwhile, intermediate effects on a given pathway are also outcomes of other processes with their own interactions and feedbacks. For example, *crop yield* occurs on numerous pathways between electricity and income but is also a product of, among other things, fertilizer use (*fertilizer*  $\rightarrow$  *crop yield*), which is an effect of income (*income*  $\rightarrow$  *fertilizer*), but which also impacts soil quality, influencing crop yields in the long run (*fertilizer*  $\rightarrow$  *soil quality*  $\rightarrow$  *crop yield*). The integrated network therefore spans temporal and spatial scales as well as causal couplings beyond the plausible bounds of any individual study.

**Figure 3** shows a small subset of the full network and demonstrates how qualitative reasoning helps interpret the structure. Here, we plot all relationships (via  $\leq 1$  intermediate concept) between *hydropower* and *biodiversity* with line widths proportional to the number of review articles describing a relationship. Relationships among the intermediate concepts are also shown. The direct relationship between hydropower and biodiversity (the ‘overall’ association) is



described by 18 articles. There are then ten other concepts which are intermediate on pathways between hydropower and biodiversity. Through qualitative reasoning, we interpret various pathways according to presumed causal ‘mechanism’. For example, hydropower limits the connectivity of freshwater aquatic habitats and changes surface water quality, implying negative biodiversity impacts. On longer timescales, hydropower displaces fossil fuels having greenhouse gas and climate benefits, which are believed to positively impact biodiversity. Forest cover emerges as a complex intermediate factor that mediates and couples other pathways. Relationships not labelled (grey lines) were determined to correspond to other, minor mechanisms. Almost all of those associations were mentioned by only one article.

Mechanistic interpretability tends to be relative to the scale and scope of individual research questions, and there is no inherent limit to the number of direct or indirect pathways possible for a given causal association. Many relationships that are described as straightforward ‘mechanisms’ in a high-level analysis might be seen as complex, multi-factor processes in more detailed studies. For example, at a high level, *surface water quality* is a ‘mechanism’ linking *hydropower* to *biodiversity* (Figure 3). However, surface water quality impacts of hydropower are themselves mediated by diverse processes including several that may also be relevant to biodiversity. Figure 4 depicts the network  $hydropower \rightarrow X \rightarrow surface\ water\ quality$ , effectively decomposing the *hydropower*  $\rightarrow$  *surface water quality* relationship from Figure 3 into more granular ‘mechanistic’ relationships. Here, we see that water quality can be impacted by (downstream) erosional and (upstream) non-

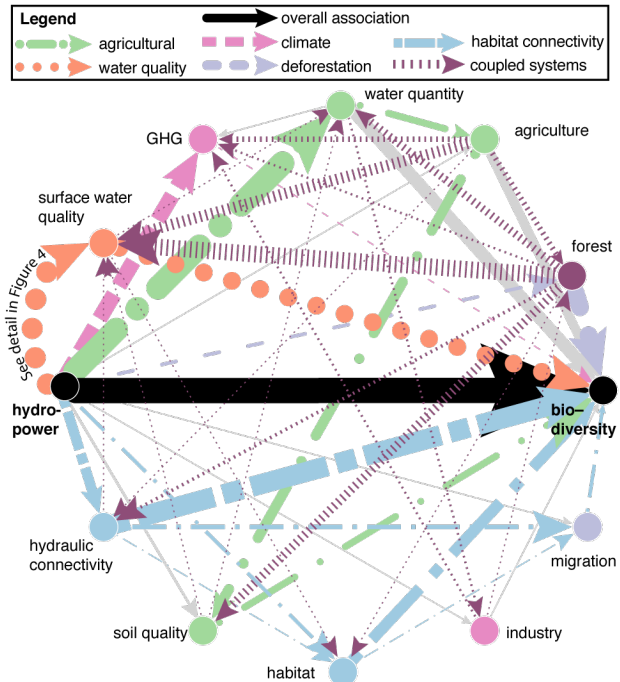


Figure 3: Sub-network from literature review showing pathways between *hydropower* and *biodiversity* (representing all pathways via one intermediate node), in addition to all relationships among intermediate concepts. Relationships with other concepts in the review are not shown. Arrow width is proportional to number of individual review articles claiming each relationship (range: 1–18). Some paths are color-coded and labeled with the presumed mechanism. Relationships with other concepts in the review are not shown. *Biodiversity* includes *aquatic life, fauna, flora* and *species*; *hydraulic conductivity* includes *river fragmentation*; *hydropower* includes *dam*; *soil quality* includes *salinity*; *water quality* includes *eutrophication* and *water temperature*; *water quantity* includes *flooding* and *river flow*.

erosional impacts on sedimentation dynamics and pollution from agriculture (promoted by increased water availability from irrigation). Hydropower reduces forest cover both directly (e.g., reservoir creation) and indirectly (via promotion of agricultural development), in turn exacerbating erosional impacts, further contributing to water quality degradation.

### 3.4. Network analysis

In the previous sections, we showed that: (i) most reviewed articles consider only simple causal pathways (**Figure 2**), (ii) most cause-and-effect relationships are not simple (**Figure 3**), and (iii) the appropriate ‘mechanistic scale’ is question-specific (**Figure 4**). Together, this suggests that selection of conceptual detail (e.g., number of pathways) is a nontrivial component of overall model development. In general, across a range of cause-and-effect associations, network complexity increases exponentially as more detailed pathways are considered (**Figure 5**). Each additional relationship allowed on a pathway introduces more concepts not only within the pathway but also across pathways in the form of mediating or compounding associations. For example, if  $A \rightarrow Y \rightarrow B$  is added to  $A \rightarrow X \rightarrow B$  as a pathway linking **A** and **B**, this could also introduce association  $Y \rightarrow X$ . This underscores the practical trade-off between consideration of mechanisms and site-specific factors on one hand, and the risk of exceeding the

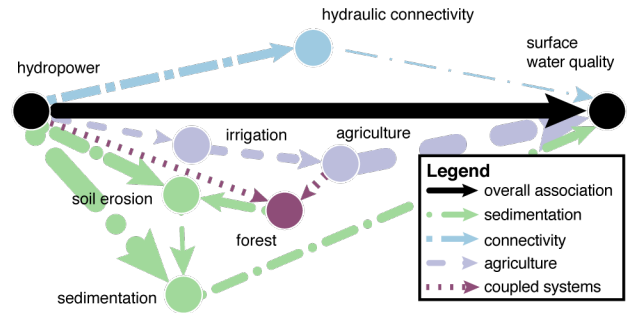


Figure 4: Sub-network showing pathways between *hydropower* and *surface water quality*, in addition to all relationships among intermediate concepts. Relationships with other concepts in the review are not shown. Arrow width is proportional to number of individual review articles claiming each relationship (range: 1–11). Some paths are color-coded and labeled with the presumed mechanism.

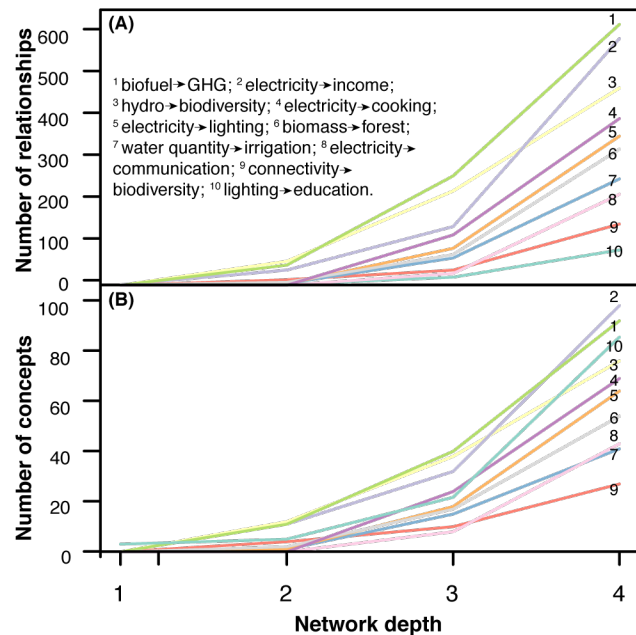


Figure 5: Comparison of the growth in complexity as a function of network depth for sub-networks linking selected cause-effect concept pairs. Depth is defined as the longest path between cause and effect for which all pathways have been identified. Complexity is indicated by the number of included (A) pairwise relationships (links) and (B) concepts (nodes). Networks at each depth also include all interactions among intermediate concepts which may incidentally introduce paths longer than the stated depth.

availability of local data needed to ultimately parameterize and evaluate the model on the other hand.

While complexity increases exponentially with network depth across a variety of cause-effect pairs, the rate of growth generally corresponds to disciplinary ‘distance’ between cause and effect. In **Figure 5**, we see that the most complex networks are those linking concepts *between* distinct disciplines (e.g., between energy sources and economic, ecological, or earth systems). The relatively less complex networks link concepts *within* disciplines (e.g., within hydrology, infrastructure, or ecology). *Lighting* → *education* is a notable exception in which deep networks have many concepts, but relatively few relationships, suggesting a limited number of causal pathways that are being studied at many different levels of mechanistic detail.

### 3.5. Evidence characterization

We evaluated the consistency of evidence in support of an association as a function of the complexity of the underlying network of relationships and found that associations connected by fewer, more directly interpretable causal mechanisms have a more consistent evidence basis than associations connected via multiple distinct causal pathways. Specifically, quantile regression results indicate that there is a greater likelihood of finding conflicting evidence on the sign of a causal relationship for high-level associations that are underlain by many different causal pathways than those that are more directly connected (**Figure 6**). For instance, while most articles associated with the relationship *hydropower* → *biodiversity* addressed the negative downstream impacts of hydropower, a subset of articles cited the potential benefits of new, productive

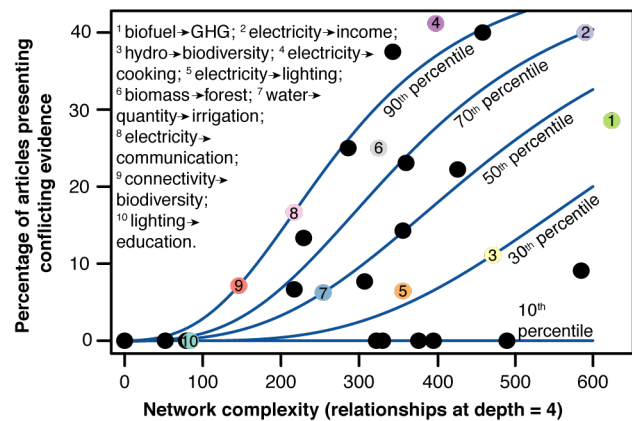


Figure 6: Percentage of review articles presenting evidence that conflicts with the prevailing direction (sign) of the cause-effect relationship, plotted against the complexity of the network linking cause to effect. Complexity is measured by the number of pairwise relationships included in networks of depth = 4. Cause-effect relationships numbered 1–10 correspond to the relationships shown in Figure 5. Lines represent estimated quantiles of a logit-like regression fit between the percentage of conflicting evidence (with a maximum of 50%) and network complexity. Quantile regression is appropriate when the predictor is hypothesised to be one of multiple possible factors constraining the value of the response. In such cases, changes in the extremes of the response distribution, rather than in the mean, are a better representation of the effect of the predictor in serving as the active limiting constraint (23). For the data shown in this figure, percentiles of 30% and greater are statistically significant ( $\alpha = 0.05$ ).

upstream ecosystems, as well as examples in which downstream impacts were nonexistent<sup>28,29</sup>. This highlights the need to evaluate interventions according to outcomes that are the direct result of mechanistically interpretable pathways rather than those that are the result of a high-level association that may yield unanticipated consequences.

We use the number of articles describing a pairwise association between concepts as a proxy for the strength of evidence supporting that association (e.g., arrow widths in **Figures 3** and **4**). The utility of this proxy is evidenced by the finding that it has a significant positive association with the total number of underlying citations mentioning the relationship (**Figure 7A**). This supports our decision to limit our literature search to review articles in order to keep our evaluation tractable. The number of articles also has a marked negative (but not statistically significant) association with the percentage of articles presenting evidence that conflicts with prevailing opinion on the sign of a causal relation (**Figure 7B**) as well as the percentage of articles presenting no evidence for the stated relation (**Figure 7C**). These associations suggest (but do not statistically establish) that evidence in support of causal relations may be ambiguous or nonexistent for inadequately studied associations, but tends to converge as research advances. Yet, it cannot be assumed that prevalence of a claimed relationship in the literature necessarily implies greater methodological rigor.

#### **4. Discussion**

Graphical networks can complement systematic reviews in supporting evidence-based practice by providing a framework for 1) understanding conflicting evidence in terms of alternative causal assumptions and 2) using accumulated evidence on socio-environmental relations to better anticipate the effects of candidate interventions. We used the impact of energy access on development, environment, and health outcomes in the developing world as a case study to show how graphical models can be used aggregate the available evidence and evaluate the evidential support for causal relations articulated at various levels of detail.

We find that relationships tend to have more consistent supporting evidence when they are more mechanistically interpretable; this analysis has suggested that ‘mechanistically interpretable’ relations are ones with relatively fewer alternate pathways between cause and effect. We have

defined ‘evidence consistency’ in terms of the proportion of published articles that concur with the prevailing sign of a causal relation. These definitions permit causal reasoning and quantitative evaluation with graphical models based on evidence originating in multiple, diverse disciplines. However, context-specific, data-driven mathematical models will still be required to quantify the actual extent to which introducing more detailed causal structures might reduce prediction error.

At a qualitative level, the network developed here can be used to identify pathways (and mediating interactions) between possible interventions and outcomes of interest. We demonstrated with the example of hydropower impacts on biodiversity (**Figure 3**) and water quality (**Figure 4**) how networks of associations drawn from the literature can be used to scope modelling and evidence needs at various scales of interest. Principal pathways and relationships can be identified by cross-referencing them with perceived importance among experts (Section 2.2) or measures of prevalence, strength, or type of evidence in the literature (Section 2.3).

There has been significant interest recently in methods for identifying ‘evidence gaps’ in the social sciences<sup>30,31</sup>. Here, we have shown how expert elicitation can be used to compare the perceived importance of a concept or relationship among experts against its prevalence in the literature. In our context, we identified several concepts and relationships that may be

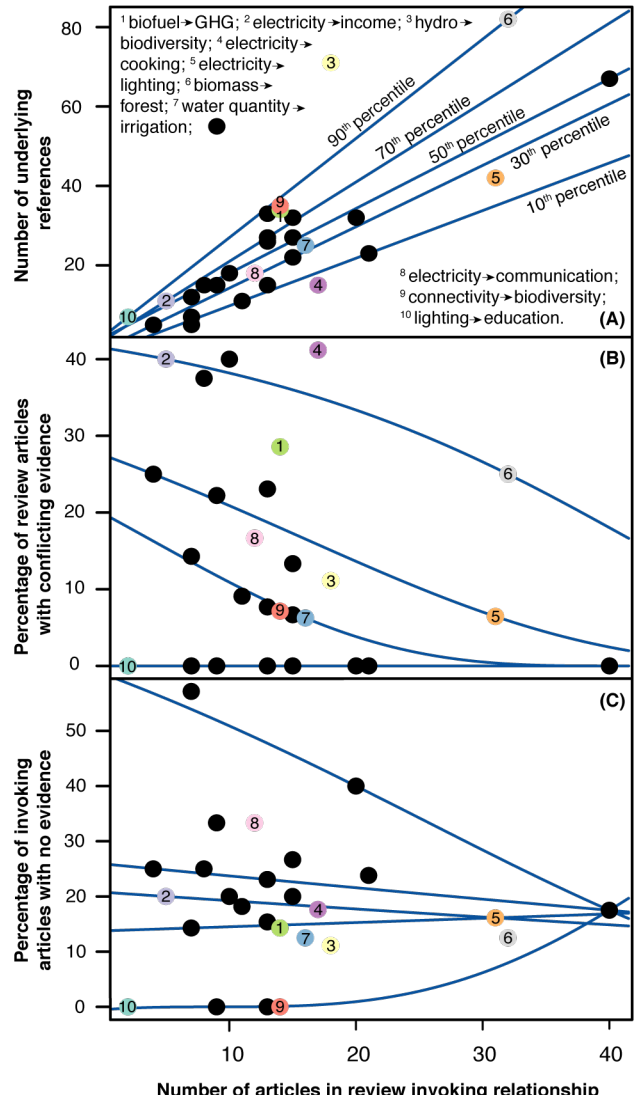


Figure 7: Selected cause-and-effect relationships plotted as number of review articles invoking each relationship against strength of evidence measured by: (A) total number of underlying references invoking each relationship; (B) percentage of review articles presenting conflicting evidence; and (C) percentage of review articles presenting no qualitative or quantitative evidence for each relationship. Cause-effect relationships numbered 1-10 correspond to those shown in Figures 5 and 6. Lines represent estimated quantiles of a linear (A) or logit-like (B and C) regression fit between the indicated quantities, with a maximum of 50% for (B) and 100% for (C). The caption for Figure 6 provides the rationale for using quantile regression to represent the relations among these variables. The quantiles shown in (A) are all statistically significant ( $\alpha = 0.05$ ), but not those shown in (B) or (C).

underrepresented in the relevant literature (e.g., income and cost effects on the uptake of energy, impacts of land rights on crop yields, effect of food preservation abilities on income, and the influence of medical facilities on the incidence of various diseases). These results however must be viewed as exploratory, as these topics were only some of many covered by the present study, and other groups of experts would likely produce different opinions on what considerations are significant.

Network analysis allows for further identification of interactions within large causal structures (as opposed to individual pairwise relationships) and may be useful for identification of interdisciplinary research needs. SI **Figure S5** identifies all relationships in this review that occur more than 10 times (top 5% of occurrence) and illustrates how commonly researched relationships connect via interlocking causal pathways. For example, the pathway: *electricity* (supply) → *pumps* → *water quantity* → *crop yield* → *income* → *electricity* (demand) demonstrates a positive feedback on electricity use enabled by the availability of water and resultant impacts on crop yield and income. Yet, no single article in the present review addressed *all* five of these relationships, and 41 of the 57 articles that addressed *any* of these five addressed only one. This example points to a need for further systems-level understanding of electricity demand and ability to pay, given that insufficient demand is a frequent barrier to rural electrification initiatives in the prevailing cost-recovery model of infrastructure investment <sup>32</sup>.

While graphical network models can complement systematic reviews, both methodologies face similar challenges. Notably, ‘mechanisms’ in the social sciences tend to be articulated in relation to specific research questions with specific temporal and spatial scales of interest <sup>2,19</sup>.

Consequently, the aggregation of ‘mechanistic’ or ‘causal’ relationships from multiple research contexts can result in highly complex structures displaying redundant pathways. Many relationships framed as a ‘fundamental mechanism’ in one context may need to be further elucidated in another. We illustrated this point using the example of *hydropower* → *surface water quality* → *biodiversity* in **Figures 3** and **4**. The result is that it is impossible for any single assessment to comprehensively account for all mediating factors.

We have also shown that decomposing a relationship into increasingly granular or mechanistically interpretable pathways increases network complexity exponentially (**Figure 5**). This increase in complexity introduces the hazard of overparameterized models in a way that

parallels the bias-variance trade-off in statistics and machine learning<sup>33</sup>. As a model of an association between two variables becomes more complex by accounting for additional variables of interest, it gains power to explain variability across diverse contexts *retrospectively* (‘lower bias’). However, arbitrarily increasing this complexity risks spurious assignment of variability to incorrect causal factors, resulting in formulations that perform worse, on average, when applied *prospectively* (‘higher variance’).

Ultimately, appropriate quantitative model formulation depends not just on causal structure but also of the availability of data for model parameterization, calibration and evaluation. Earlier, we reviewed the relatively simple causal structures supported by the existing evidence base (**Figure 2**). This evidence base will most likely only support the evaluation of interventions with impacts along well-studied causal chains or that are causally proximal. Of course, sometimes assumptions can be made to constrain model complexity without sacrificing predictive power. For example, as noted previously, a site-specific analysis of the biodiversity impacts of hydropower may not need to consider long-term global biodiversity benefits related to transitions away from fossil fuels, or may not be concerned with forest ecosystems specifically. Thus, suppressing ‘greenhouse gases’ or ‘forest’ from the pathway between hydropower and biodiversity (**Figure 3**) can reduce the number of network relations by a third (from 53 to 36).

Like systematic reviews, our graphical model summarizes causal relations reported in the literature and therefore reflects biases of past research. For example, a large proportion of the research we reviewed characterized the impacts of biofuel use, cook stove technologies and indoor air quality impacts (**Figure 1**). Previous reviews have suggested that the focus on health impacts of indoor cooking transitions may be disproportionate to its importance<sup>34,35</sup>. Additionally, we faced a trade-off between respecting the original language used by authors and harmonizing terminology for similar concepts. Precise usage for commonly terms such as ‘biofuel’ and ‘electrification’ varies. There is also no consensus on what level of availability, affordability or reliability of electricity constitutes ‘access’<sup>36</sup>. We tended to use broad, inclusive terms to aggregate similar concepts to minimize the number of nodes and relationships in the overall network and more fully characterize the space of possible causal relations. More focused networks may need more refined definitions of some concepts.

Overall, our analysis supports the utility of graphical models for organizing evidence across highly interdisciplinary communities of practice and for reconceptualizing studied associations as the result of interacting processes. Graphical models increase the transparency of subjective modelling decisions and bring into relief the inherent complexities of socio-environmental and biophysical systems. Models developed from large literature reviews such as that presented here can reveal previously overlooked associations, interactions, and causal pathways. Yet, model conceptualization has an irreducibly subjective dimension, and any real socio-environmental system is likely to defy description by a single graphical structure. In future work, we plan to concentrate on more focused associations (notably the benefits of rural electrification on income) to consider the challenges posed by interdisciplinary models to quantitative parameterization.

## **5. Acknowledgments**

We thank Elizabeth Beyer (North Carolina State University) for help organizing results of the literature search and Spencer Perkins (Duke University) for help coding the results of the expert elicitation survey. We gratefully acknowledge funding from the Duke University Nicholas Institute for Environmental Policy Solutions Catalyst Program (2017-18 and 2018-19) and the Margaret A. Cargill Philanthropies. We thank our colleagues in The Bridge Collaborative, a collaboration between Duke University, the International Food Policy Research Institute, PATH and The Nature Conservancy, for their cooperation and support.



## 6. References

- 1 Boaz, A., Ashby, D. & Young, K. Systematic Reviews: What have they got to offer evidence based policy and practice? (London, UK: ESRC UK Centre for Evidence Based Policy and Practice, Queen Mary University of London, 2002).
- 2 Clegg, S. Evidence-based practice in educational research: a critical realist critique of systematic review. *Brit J Sociol Educ* **26**, 415-428, doi: 10.1080/01425690500128932 (2005).
- 3 Grant, M. J. & Booth, A. A Typology of reviews: an analysis of 14 review types and associated methodologies. *Health Info Libr J* **26**, 91-108, doi: 10.1111/j.1471-1842.2009.00848.x (2009).
- 4 Pullin, A. S. & Knight, T. M. Doing more good than harm – Building an evidence-base for conservation and environmental management. *Biol Conserv* **142**, 931-934, doi: 10.1016/j.biocon.2009.01.010 (2009).
- 5 Davies, H. T. O., Nutley, S. M. & Smith, P. C. in *What works? Evidence-based policy and practice in public services* (eds Huw T. O. Davies, Sandra M. Nutley, & Peter C. Smith) (The Policy Press, 2000).
- 6 Davies, H. T. O. & Nutley, S. M. *Healthcare: evidence to the fore*. (The Policy Press, 2000).
- 7 Kleinman, A. in *Companion Encyclopedia of the History of Medicine* (eds W. F. Bynum & Roy Porter) (Routledge, 1993).
- 8 White, H. & Waddington, H. Why do we care about evidence synthesis? An introduction to the special issue on systematic reviews. *J Dev Effect* **4**, 351-358, doi: 10.1080/19439342.2012.711343 (2012).
- 9 Proctor, S. Linking philosophy and method in the research process: the case for realism. *Nurs Res* **5**, 73-90 (1998).
- 10 Mallett, R., Hagen-Zanker, J., Slater, R. & Duvendack, M. The benefits and challenges of using systematic reviews in international development research. *J Dev Effect* **4**, 445-455, doi: 10.1080/19439342.2012.711342 (2012).
- 11 Langer, L. & Stewart, R. What have we learned from the application of systematic review methodology in international development? – a thematic overview. *J Dev Effect* **6**, 236-248, doi: 10.1080/19439342.2014.919013 (2014).
- 12 Tallis, H., Merz, B. J., Huang, C., Kreis, K., Olander, L. & Ringler, C. A Call to Action for Health, Environment and Development Leaders: A Report by the Bridge Collaborative. (Washington, D.C.: The Nature Conservancy, 2017).
- 13 Little, J. C., Hester, E. T., Elsworth, S., Filz, G. M., Sandu, A., Carey, C. C., Iwanaga, T. & Jakeman, A. J. A tiered, system-of-systems modeling framework for resolving complex socio-environmental policy issues. *Environ Modell Softw* **112**, 82-94, doi: 10.1016/j.envsoft.2018.11.011 (2019).
- 14 Rosa, E. A. & Dietz, T. Climate change and society: Speculation, construction and scientific investigation. *Int Sociol* **13**, 421-455 (1998).

- 15 Hartvigsson, E., Ahlgren, E. O. & Molander, S. Tackling complexity and problem formulation in rural electrification through conceptual modelling in system dynamics. *Syst Res Behav Sci*, 1-13, doi: 10.1002/sres.2611 (2019).
- 16 Carriger, J. F., Dyson, B. E. & Benson, W. H. Representing causal knowledge in environmental policy interventions: Advantages and opportunities for qualitative influence diagram applications. *Integr Environ Assess Manag* **14**, 381-394, doi: 10.1002/ieam.2027 (2018).
- 17 Tallis, H., Kreis, K., Olander, L., Ringler, C., Ameyaw, D., Borsuk, M. E., Fletschner, D., Game, E., Gilligan, D. O., Jeuland, M., Kennedy, G., Masuda, Y. J., Mehta, S., Miller, N., Parker, M., Pollino, C. A., Rajaratnam, J., Wilkie, D. & Zhang, W. Bridge Collaborative Practitioner's Guide. (Washington, D.C.: The Nature Conservancy, 2017).
- 18 Centers for Disease Control and Prevention (CDC) Division for Heart Disease and Stroke Prevention. Evaluation Guide: Developing and Using a Logic Model. (Washington, D.C.: 2010).
- 19 Cornish, F. Evidence synthesis in international development: a critique of systematic reviews and a pragmatist alternative. *Anthropol Med* **22**, 263-277, doi: 10.1080/13648470.2015.1077199 (2015).
- 20 Mayntz, R. Mechanisms in the analysis of social macro-phenomena. *Philos Soc Sci* **34**, 237-259, doi: 10.1177/0048393103262552 (2016).
- 21 Rittel, H. W. & Webber, M. M. Dilemmas in a general theory of planning. *Policy Sci* **4**, 155-169 (1973).
- 22 Liu, J., Dietz, T., Carpenter, S. R., Alberti, M., Folke, C., Moran, E., Pell, A. N., Deadman, P., Kratz, T., Lubchenco, J., Ostrom, E., Ouyang, Z., Provencher, W., Redman, C. L., Schneider, S. H. & Taylor, W. W. Complexity of coupled human and natural systems. *Science* **317**, 1513-1516, doi: 10.1126/science.1144004 (2007).
- 23 Mendeley Ltd. Mendeley Desktop V.1.17.11. (Amsterdam, Netherlands: Elsevier, 2017).
- 24 R Core Team. R: A language and environment for statistical computing. (Vienna, Austria: R Foundation for Statistical Computing, 2018).
- 25 Müller, K., Wickham, H., James, D. A. & Falcon, S. RSQLite: SQLite interface for R. (2011).
- 26 Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal, Complex Systems* **1695**, 1-9 (2006).
- 27 Cade, B. S., Terrell, J. W. & Schroeder, R. L. Estimating effects of limiting factors with regression quantiles. *Ecology* **80**, 311-323, doi: 10.1890/0012-9658(1999)080[0311:Eeolfw]2.0.Co;2 (1999).
- 28 Li, Z., He, D. & Feng, Y. Regional hydropolitics of the transboundary impacts of the Lancang cascade dams. *Water Int* **36**, 328-339, doi: 10.1080/02508060.2011.585447 (2011).
- 29 Yüksel, I. Hydropower for sustainable water and energy development. *Renew Sust Energ Rev* **14**, 462-469, doi: 10.1016/j.rser.2009.07.025 (2010).

- 30 Snilstveit, B., Bhatia, R., Rankin, K. & Leach, B. 3ie evidence gap maps: a starting point for strategic evidence production and use. (New Delhi, India: International Initiative for Impact Evaluation (3ie), 2017).
- 31 Müller-Bloch, C. & Kranz, J. in *International Conference on Information Systems*.
- 32 Cook, P. Infrastructure, rural electrification and development. *Energy Sustain Dev* **15**, 304-313, doi: 10.1016/j.esd.2011.07.008 (2011).
- 33 Geurts, P. in *Data Mining and Knowledge Discovery Handbook* (eds Oded Maimon & Lior Rokach) 733-746 (Springer US, 2010).
- 34 Jeuland, M., Fetter, T. R., Li, Y., Pattanayak, S. K. & Usmani, F. Is energy a golden thread? A systematic review of the impacts of modern and traditional energy use in low- and middle-income countries. (under review).
- 35 Usmani, F. Energy as the Golden Thread: What Do We Know? (Durham, NC: Duke University Energy Access Project, 2018).
- 36 Pachauri, S. Reaching an international consensus on defining modern energy access. *Curr Opin Env Sust* **3**, 235-240, doi: 10.1016/j.cosust.2011.07.005 (2011).

## Supplemental Information

### Graphical models and the challenge of evidence-based practice in development and sustainability

Ryan S.D. Calder<sup>a\*</sup>, Andrea Alatorre<sup>b</sup>, Rebecca S. Marx<sup>b</sup>, Varun Mallampalli<sup>a</sup>, Sara A. Mason<sup>c</sup>, Lydia P. Olander<sup>c</sup>, Marc Jeuland<sup>d</sup>, Mark E. Borsuk<sup>a</sup>

<sup>a</sup> Department of Civil and Environmental Engineering, Pratt School of Engineering, Duke University, Durham NC 27708

<sup>b</sup> Nicholas School of the Environment, Duke University, Durham NC 27708

<sup>c</sup> Nicholas Institute for Environmental Policy Solutions, Duke University, Durham NC 27708

<sup>d</sup> Sanford School of Public Policy, Duke University, Durham NC 27708

\* Corresponding author: 1116 Hudson Hall, Box 90287; email: ryan.calder@duke.edu; phone: (919) 660-6883; fax: (919) 660-5219

### Contents

Supplemental methods.....	1
1. Literature review .....	1
1.1. Scope of study and initial search .....	1
1.2. Selection of articles.....	1
1.3. Coding and review of articles.....	1
1.4. Elucidation of network and formulation of secondary search .....	1
1.5. Evaluation of evidence underlying key connections.....	2
1.6. Statistical analysis of evidence .....	2
2. Expert elicitation workshops.....	7
2.1. Participant recruitment.....	7
2.2. Elicitation interviews .....	7
Supplemental figures and tables from main text .....	11
Appendix A: Web of Science search syntax .....	17
Appendix B: Consent materials for expert elicitation workshops.....	23
References for Supplemental Information .....	32

#### List of figures from supplemental methods

Figure SM1: Article selection tree.....	6
Figure SM2: Sample expert elicitation network diagrams.....	10
Figure SM3: Example influence diagram .....	27

#### List of tables from supplemental methods

Table SM1: Demographic information of participants.....	8
Table SM2: Professional information for participants.....	9

#### List of supplemental figures from main text

Figure S1: Occurrence of country names among article abstracts.....	11
Figure S2: Concepts supplied to experts in elicitation workshops .....	12
Figure S3: Relations drawn by experts from predefined concepts .....	13
Figure S4: Relational network derived from guided expert elicitation exercises.....	14
Figure S5: Relations mentioned in more than 10 articles (top 5% of occurrence).....	15

#### List of supplemental tables from main text

Table S1: Summary of evidence characterization .....	16
--	----

## Supplemental methods

### 1. Literature review

#### 1.1. Scope of study and initial search

We used Web of Science (WoS) to conduct a structured search of the peer-reviewed literature for review articles describing links between energy access and health, environmental or development outcomes in the developing world. The search syntax is included in **Appendix A**. Using Boolean logic, we searched for articles whose WoS *topic* field included at least one keyword from the “energy access” (e.g., “electri\* access”) set, and at least one keyword from at least one (but not necessarily more than one) of the “environment” (e.g., “biodiversity”), “development” (e.g., “income”) and “health” (e.g., “mortality”) sets. We also required the *topic* field to contain at least one out of a list of “developing countries” keywords consisting of a list of developing countries and variations and synonyms of “developing countries”. We confined this search to review articles and excluded irrelevant subjects. All searches were performed in 2017-18. Only articles published in English were considered. To evaluate possible bias due to this review article constraint, we compared representation of concepts in our review to the breadth of evidence in underlying primary research and report this evaluation as part of our evidence review in the main text.

#### 1.2. Selection of articles

The initial search yielded 1,429 unique results. Articles were added to Mendeley (Mendeley Ltd. 2017) where they were sorted and reviewed. Coauthors (AA and RM) and a research assistant (EB) reviewed abstracts for relevance and excluded articles that did not pertain to energy access or the outcomes of interest (138 articles), presented only a focused technical analysis (350 articles), discussed only energy goals with no discussion of causal associations (310 articles), discussed only climate impacts (21 articles) or focused solely on governance (256 articles). Following this process of article selection, 354 articles (25%) were retained for coding and review. **Figure SM1** at the end of Section 1 presents an article selection tree.

#### 1.3. Coding and review of articles

Coauthors (AA and RM) and a research assistant (EB) used the abstracts of the 354 articles retained for analysis from the initial search to record the relationships claimed or evaluated by each article. Of these, 140 (40%) were subsequently reviewed in full either because causal connections claimed or studied could not be determined from the abstract (100 articles) or because these same articles were returned in a further secondary search described below (21 articles). Relationships were recorded as Mendeley “tags” in the format  $A \rightarrow B$  where “A” is a cause and “B” is effect, according to the terminology used by the authors of each article. After all articles were coded, coauthors (AA, RM, RSDC and SAM) reviewed the coding and consolidated synonyms and common concepts across articles (e.g., “picogrid”, “nanogrid”, “microgrid” consolidated under “small grid”). All tags were saved into Mendeley metadata.

#### 1.4. Elucidation of network and formulation of secondary search

All authors met to discuss the structure of the relational network and identified a number of cause-effect pairs lacking a mechanistically interpretable pathway. For example, forests mediate numerous biophysical phenomena, each of which are potentially important to biodiversity outcomes and the overall relationship forest  $\rightarrow$  biodiversity was insufficient for us to outline a

pathway. A secondary search targeting these associations was carried out to further elucidate those pathways. These searches are summarized in **Table SM1** at the end of Section 1. Authors carried out a WoS search followed by selection, reviewing, and coding of articles (described in **Sections 1.2–0**) for review articles returned by the secondary search. All of these articles were reviewed in full to elucidate underlying mechanistic pathways. This review retained 232 unique articles, including 21 that had been identified in the first search. In total, across the first and second searches, 561 unique articles were identified for further review. **Figure SM1** at the end of Section 1 presents an article selection tree.

### 1.5. Evaluation of evidence underlying key connections

Tallis et al. (2017) propose a rubric to evaluate evidence underlying claimed relationships in terms of (1) number of types of evidence (e.g., longitudinal, qualitative); (2) consistency of results; (3) the extent to which the methods are documented and accepted; and (4) the applicability of evidence to the question or case at hand. This rubric aims to accommodate the different standards and best practices for study design across sectors and to better reflect the state of knowledge across highly complex and interdisciplinary relational spaces.

We evaluated a subset of the relationships identified here according to an adapted version of this rubric. For (3), instead of expressing our subjective judgment about the acceptability of methods used in available evidence, we record what these methods were among (i) expert knowledge or case report; (ii) quantitative longitudinal study; (iii) quantitative correlative study; (iv) inferences from quantitative modeling; (v) inferences from qualitative theory; and (vi) conjecture with no apparent evidence. This information was derived directly from review articles and, where necessary, by consulting underlying primary articles. We do not evaluate relationships according to criterion (4) because applicability of a relationship is a site-specific judgment. Coauthors AT and RM evaluated saved articles for a subset of the links. The relationships selected for evidence characterization correspond generally to the most commonly identified relationships in the review (i.e., a subset of those presented in main text **Figure 1**) but the characterization process also aimed to cover a diversity of time- and length-scales, and both high-level statistical associations (e.g., electricity  $\rightarrow$  education) and mechanistically interpretable physical processes (e.g., damming  $\rightarrow$  hydraulic connectivity). In total, given time resource constraints, 29 relationships were retained for this detailed evidence characterization.

### 1.6. Statistical analysis of evidence

We used quantile regression to analyze the associations among evidence variables shown in Figures 6 and 7. Quantile regression is an appropriate statistical modelling method when the predictor variable is hypothesized to be one of multiple possible factors acting to constrain the value of the response variable (Cade et al. 1999).

Figure 6 relates the percentage of review articles presenting evidence that conflicts with the prevailing sign of selected cause-and-effect relationships (response variable,  $p$ ) to the complexity of the network linking cause to effect (predictor variable,  $x$ ), where complexity is measured by the number of pairwise relationships (links) included in networks of depth = 4. We interpret the total number of relationships in a network as theoretically constraining the opportunities for the sign of an association to differ from the prevailing evidence. Other factors (e.g., experimental design details, site-specific applicability of relations, or paucity of studies) may also limit the frequency of conflicting research results. Indeed, subsequent analysis (see below) reveals that the

total number of articles describing a relationship may also constrain the frequency of conflicting evidence.

To perform the quantile regression, we used the *quantreg* package in R. Because the values of the response variable  $p$  are constrained by definition to the interval  $[0, 0.5]$ , we first transformed  $p$  using a modified logit transformation, replacing the typical logarithmic function with the cube root to accommodate values of 0 and replacing the typical upper limit of 1 with 0.5. This yields:

$$y = \sqrt[3]{\left(\frac{p}{0.5-p}\right)} = \beta_0 + \beta_1 x \quad (1)$$

where  $\beta_0$  and  $\beta_1$  are the slope and intercept of the linearized model. The R function calls and output for a quantile regression (using a bootstrap method to estimate standard errors) for the 10<sup>th</sup>, 30<sup>th</sup>, 50<sup>th</sup>, 70<sup>th</sup>, and 90<sup>th</sup> percentiles are as follows:

```
> y = (p/(.5-p))^(1/3)
> rqfit = rq(y ~ x, tau = c(0.1, 0.3, 0.5, 0.7, 0.9))
> summary(rqfit, se = "boot")
```

tau: [1] 0.1  
Coefficients:

	Value	Std. Error	t value	Pr(> t )
(Intercept)	0.00000	0.08286	0.00000	1.00000
x	0.00000	0.00044	0.00000	1.00000

tau: [1] 0.3  
Coefficients:

	Value	Std. Error	t value	Pr(> t )
(Intercept)	-0.13442	0.15190	-0.88490	0.38402
x	0.00168	0.00075	2.24609	0.03308

tau: [1] 0.5  
Coefficients:

	Value	Std. Error	t value	Pr(> t )
(Intercept)	-0.00206	0.20241	-0.01017	0.99196
x	0.00206	0.00069	2.99290	0.00585

tau: [1] 0.7  
Coefficients:

	Value	Std. Error	t value	Pr(> t )
(Intercept)	-0.00270	0.19762	-0.01364	0.98922
x	0.00270	0.00077	3.48953	0.00168

tau: [1] 0.9  
Coefficients:

	Value	Std. Error	t value	Pr(> t )
(Intercept)	0.04642	0.23157	0.20047	0.84262
x	0.00343	0.00091	3.78050	0.00079

After parameter estimation, we back-transform eq. (1) to plot quantile curves in the original metric of the response variable as:

$$p = \frac{0.5(\beta_0 + \beta_1 x)^3}{1 + (\beta_0 + \beta_1 x)^3} \quad (2)$$

Quantile regression was performed similarly for Figure 7(C), with the same response variable  $p$ , but with the number of review articles found that describe each relationship serving as the predictor variable  $z$ . In this case, the R function calls and output for the 10<sup>th</sup>, 30<sup>th</sup>, 50<sup>th</sup>, 70<sup>th</sup>, and 90<sup>th</sup> percentiles are as follows:

```
> y = (p/(.5-p))^(1/3)
> rqfit = rq(y ~ z, tau = c(0.1, 0.3, 0.5, 0.7, 0.9))
> summary(rqfit, se = "boot")
```

```
tau: [1] 0.1
Coefficients:
      Value Std. Error t value Pr(>|t|)
(Intercept) 0.00000 0.06177  0.00000 1.00000
x            0.00000 0.00043  0.00000 1.00000
```

```
tau: [1] 0.3
Coefficients:
      Value Std. Error t value Pr(>|t|)
(Intercept) -0.13442 0.09491  -1.41628 0.16813
x            0.00168 0.00076   2.21890 0.03509
```

```
tau: [1] 0.5
Coefficients:
      Value Std. Error t value Pr(>|t|)
(Intercept) -0.00206 0.19048  -0.01081 0.99146
x            0.00206 0.00069   2.99935 0.00575
```

```
tau: [1] 0.7
Coefficients:
      Value Std. Error t value Pr(>|t|)
(Intercept) -0.00270 0.18424  -0.01463 0.98844
x            0.00270 0.00074   3.62847 0.00117
```

```
tau: [1] 0.9
Coefficients:
      Value Std. Error t value Pr(>|t|)
(Intercept) 0.04642 0.20117   0.23077 0.81923
x            0.00343 0.00092   3.72492 0.00091
```

In Figure 7(B), the response variable is the percentage of reviewed articles presenting no evidence for each relationship  $q$ , constrained in this case to the range  $[0,1]$ .

```
> y = (q/(1-q))^(1/3)
> rqfit = rq(y ~ z, tau = c(0.1, 0.3, 0.5, 0.7, 0.9))
> summary(rqfit, se = "boot")
```

```
tau: [1] 0.1
Coefficients:
      Value Std. Error t value Pr(>|t|)
(Intercept) 0.00000 0.02671  0.00000 1.00000
z            0.00000 0.00267  0.00000 1.00000
```

```
tau: [1] 0.3
Coefficients:
      Value Std. Error t value Pr(>|t|)
(Intercept) 0.00000 0.32010  0.00000 1.00000
z            0.00000 0.01412  0.00000 1.00000
```

```
tau: [1] 0.5
Coefficients:
      Value Std. Error t value Pr(>|t|)
(Intercept) 0.66298 0.23582   2.81141 0.00907
z           -0.01609 0.01381  -1.16544 0.25403
```

```
tau: [1] 0.7
Coefficients:
      Value Std. Error t value Pr(>|t|)
(Intercept) 0.73533 0.15390   4.77796 0.00006
```



```
z          -0.01049  0.00969  -1.08270  0.28851
```

```
tau: [1] 0.9
```

```
Coefficients:
```

	Value	Std. Error	t value	Pr(> t )
(Intercept)	0.90695	0.10895	8.32466	0.00000
z	-0.00667	0.00908	-0.73524	0.46853

Finally, in Figure 7(A) the total number of underlying references corroborating each relationship  $r$  was used as the response variable. As this variable is not a percentage and is therefore not constrained, it was used directly as the untransformed response variable  $y$  of the linear function in eq. (1).

```
> rqfit = rq(r ~ z, tau = c(0.1, 0.3, 0.5, 0.7, 0.9))  
> summary(rqfit, se = "boot")
```

```
tau: [1] 0.1
```

```
Coefficients:
```

	Value	Std. Error	t value	Pr(> t )
(Intercept)	-2.20000	3.56984	-0.61627	0.54287
z	1.20000	0.34756	3.45264	0.00184

```
tau: [1] 0.3
```

```
Coefficients:
```

	Value	Std. Error	t value	Pr(> t )
(Intercept)	-1.18182	3.55212	-0.33271	0.74192
z	1.54545	0.31565	4.89614	0.00004

```
tau: [1] 0.5
```

```
Coefficients:
```

	Value	Std. Error	t value	Pr(> t )
(Intercept)	1.66667	4.54977	0.36632	0.71698
z	1.63333	0.45682	3.57543	0.00134

```
tau: [1] 0.7
```

```
Coefficients:
```

	Value	Std. Error	t value	Pr(> t )
(Intercept)	1.00000	6.35574	0.15734	0.87615
z	2.00000	0.54704	3.65602	0.00109

```
tau: [1] 0.9
```

```
Coefficients:
```

	Value	Std. Error	t value	Pr(> t )
(Intercept)	2.00000	18.87576	0.10596	0.91640
z	2.50000	0.99091	2.52294	0.01783

Table SM1: Secondary searches run after development of preliminary network

Secondary search (cause → effect) <sup>a</sup>	Articles returned <sup>b</sup>	Articles retained (%)
air quality → biodiversity	21	10 (48%)
eutrophication → nutrition	23	5 (22%)
food waste → nutrition	10	4 (40%)
forest → biodiversity	26	13 (50%)
irrigation → crop yield	48	30 (63%)
lighting or electricity → education	75	25 (33%)
lighting or electricity → employment	171	59 (35%)
river fragmentation → biodiversity	2	1 (50%)
soil quality → biodiversity	45	20 (44%)
soil quality → crop yield	48	24 (50%)
temperature control → nutrition	48	7 (15%)
water quality → biodiversity	151	52 (34%)
Unique articles <sup>c</sup>	633	232

<sup>a</sup> See Appendix A for exact search pattern used.

<sup>b</sup> Unique articles (excluding duplicates).

<sup>c</sup> Less than sum of individual searches because some articles were returned in more than one search.

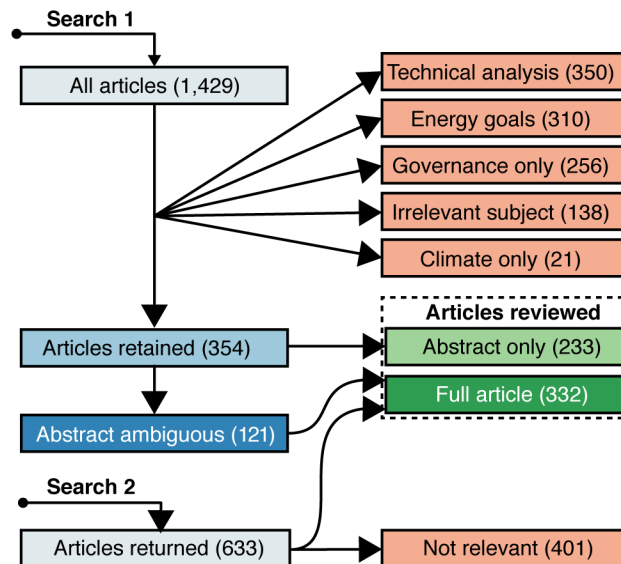


Figure SM1: Article selection tree showing classification of articles returned in two Web of Science searches. Search 1 yielded 1,494 articles from which 65 duplicates were removed yielding 1,429 unique articles. Rejected articles in Search 2 were not further sorted. The 354 articles retained following Search 1 include 21 articles that were also returned by Search 2 and hence reviewed in full.

## 2. Expert elicitation workshops

### 2.1. Participant recruitment

The Sustainable Energy Transitions Initiative (SETI) at Duke University holds an annual conference of scholars and practitioners working on energy transitions in the developing world. We invited attendees to participate in voluntary expert elicitation interviews that we held to characterize expert belief about the structure of interventions, outcomes and intermediate/confounding effects relevant to global energy transitions. We emailed 64 conference registrants (all expected attendees excluding those affiliated with Duke University) of whom 30 (47%) agreed to participate in the interviews held over May 15–17, 2018.

Participants were predominantly academics (80%) with 1–10 years of experience (83%) representing institutions in North America (60%). Of the professionals from North American institutions, 40% had extensive personal and professional connections to developing countries in Africa and South Asia. **Table SM2** and **Table SM3** at the end of Section 2 report the distributions of participant demographic and professional information.

### 2.2. Elicitation interviews

Interviews were carried out on an individual basis and were performed by coauthors AA, RC and VM using a common script. Interviews began with recording participants' informed consent, an explanation of the intent of the research, and an introduction of the concept of a results chain using a figure based on an example from Morgan et al. (2002). **Appendix B** contains all interview recruitment and consent materials, the interview script, and approval of our research protocol from the Duke University Institutional Review Board (Protocol 2018-0497). The figure used to explain causal diagrams to participants is included in Appendix B as Figure SM3.

Participants completed a “guided” exercise as described in the main text, assembling a common list of ~70 predefined concepts on a whiteboard and drawing causal connections in dry-erase marker. As of the first round of the literature review, these concepts were the most widely identified and interconnected. Concepts were supplied on printed readhesive labels, and each participant received a fresh batch of labels to avoid inter-participant biasing.

To evaluate the likely familiarity of experts with the concepts they arranged in causal networks, we also asked participants to participate in an “unguided” exercise. This was carried out before the guided exercise described above. Here, participants drew on a whiteboard the causal structure of interventions, intermediate phenomena and outcomes of interest in their individual realms of practice or research. In these unguided exercises, experts identified a total of 295 unique concepts across all individual interviews, of which 30 were also present in the guided exercise (predefined concepts provided to all experts). Therefore, about 45% of the terms supplied in the guided exercise were represented across all unguided exercises. Overall, there is a positive and significant ( $p < 0.05$ ) correlation between frequency of invocation of concepts in the guided and unguided exercises ( $R^2 = 0.17$ ). Of the remaining concepts that appeared in the unguided exercise, roughly 90% appeared in only one expert's unguided network.

Interviewers recorded the audio from all interviews and photographed the causal networks assembled. **Figure SM2** at the end of Section 2 displays sample causal diagrams from one interview. The networks were coded by a research assistant (SP).

Table SM2: Demographic information of participants (one response possible)

Age range	Fraction	n
18 – 24	0%	0
25 – 34	60%	18
35 – 44	33%	10
45 – 54	3%	1
55 – 64	3%	1
≥ 65	0%	0
Total (N)	100%	30
Gender		
Male	67%	20
Female	33%	10
Total (N)	100%	30
Years of experience in reported sector <sup>a</sup> (see Table SM3)		
< 1 year	0%	0
1 – 5 years	40%	12
6 – 10 years	43%	13
11 – 15 years	13%	4
16 – 20 years	3%	1
> 20 years	0%	0
Total (N)	100%	30
Location of home institution <sup>b</sup>		
United States/Canada	60%	18
Europe	20%	6
Asia	10%	3
Africa	7%	2
South America	3%	1
Total (N)	100%	30

<sup>a</sup> See Table SM3 for reported sectors of practice.

<sup>b</sup> Based on registration data.

Table SM3: Professional information for participants (multiple responses possible)

Sector	Fraction	n
Academia/other research	80%	24
NGO	20%	6
Government	3%	1
Private sector	7%	2
Other <sup>a</sup>	3%	1
Activities constituting the majority of work time among participants		
Investment analysis or financial planning	0%	0
Communications	7%	2
Policy analysis/non-academic writing	23%	7
Policy analysis or development	23%	7
Research, data analysis or writing academic papers	87%	26
Project management	23%	7
Teaching students or training staff	30%	9
Proposal writing	10%	3
Other <sup>b</sup>	10%	3
How participants learn of new evidence in their sector		
Academic research papers	93%	28
Advertisements	3%	1
Conferences/networking	93%	28
Mass media	33%	10
Reports or white papers	60%	18
Social media	40%	12
Trade magazines	3%	1
Word of mouth	33%	10
Other <sup>c</sup>	3%	1

<sup>a</sup> Other answer supplied: “Intergovernmental organization ([Redacted name of organization])”.

<sup>b</sup> Other answers supplied: “Energy assessment and modeling”; “Consulting”; “Fundraising, executive duties, advocacy”.

<sup>c</sup> Other answer supplied: “Government notifications/regulatory orders”.

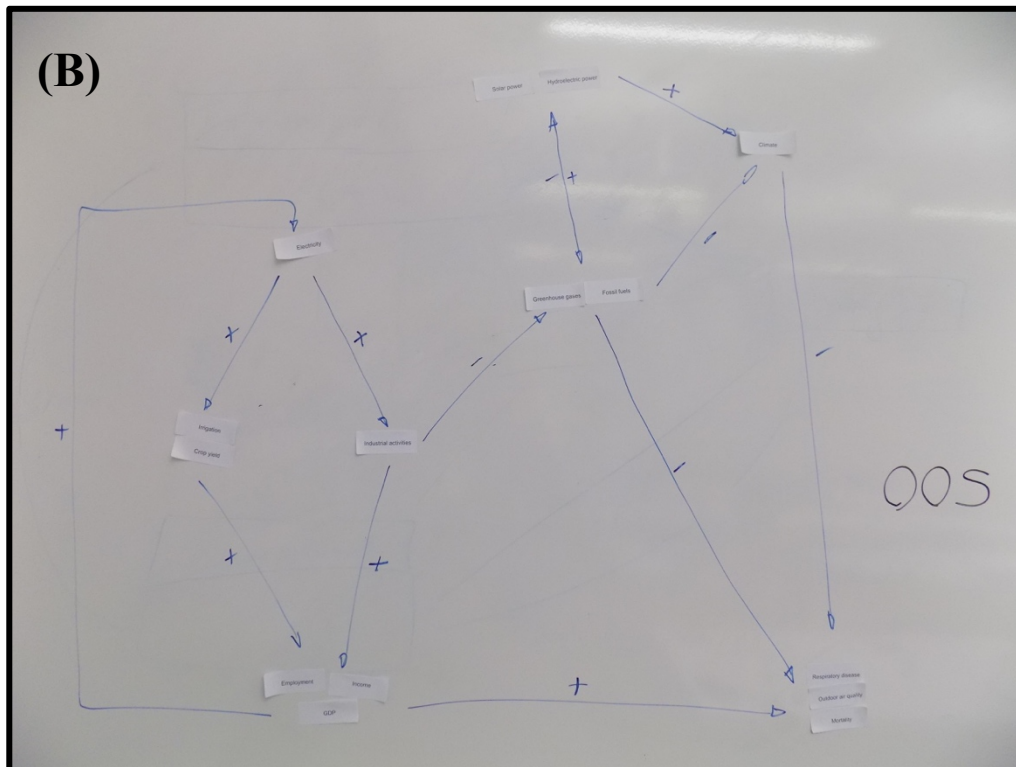
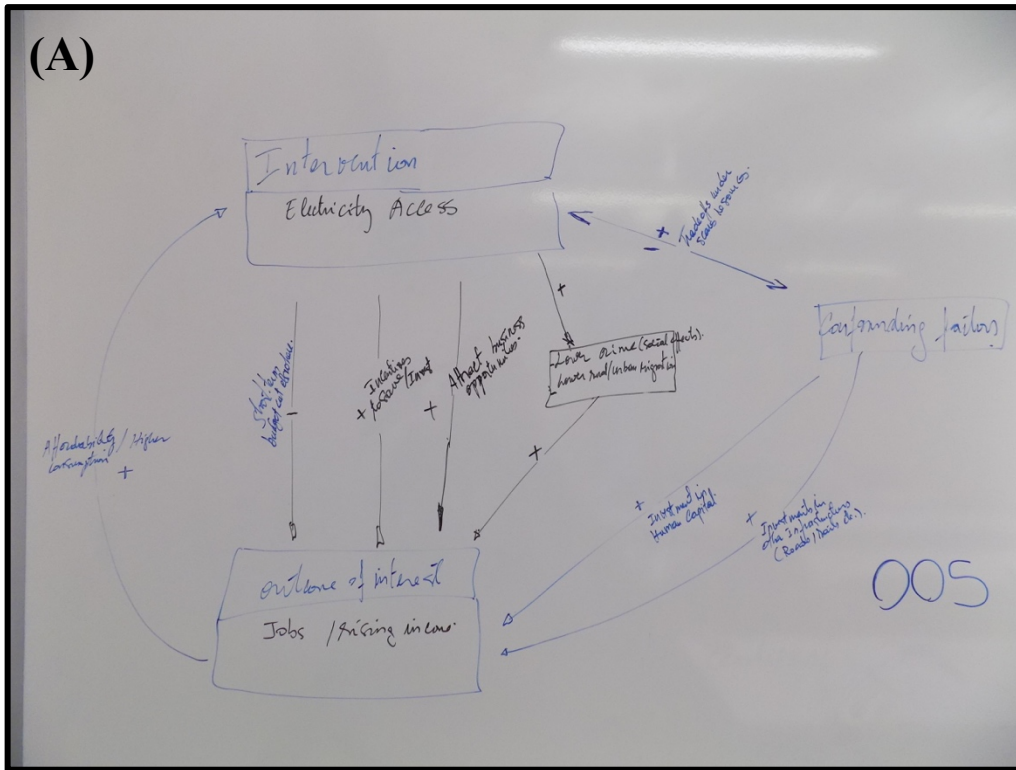


Figure SM2: Sample expert elicitation network diagrams from unguided (A) and guided (B) exercises.

### Supplemental figures and tables from main text

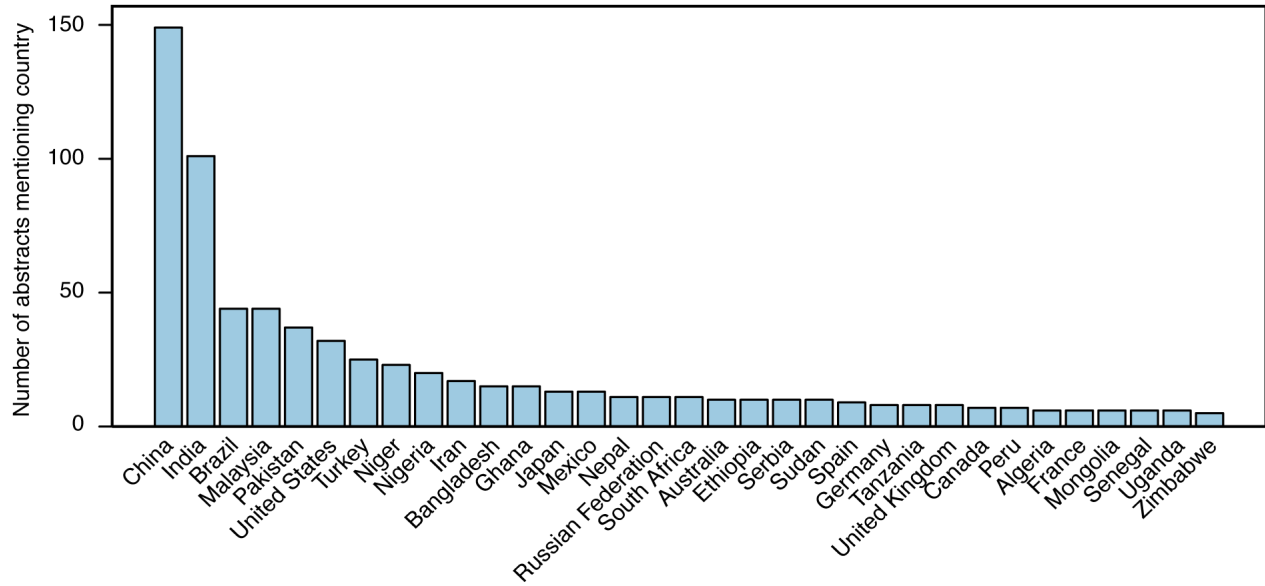


Figure S1: Occurrence of country names among article abstracts (including commonly used variants)

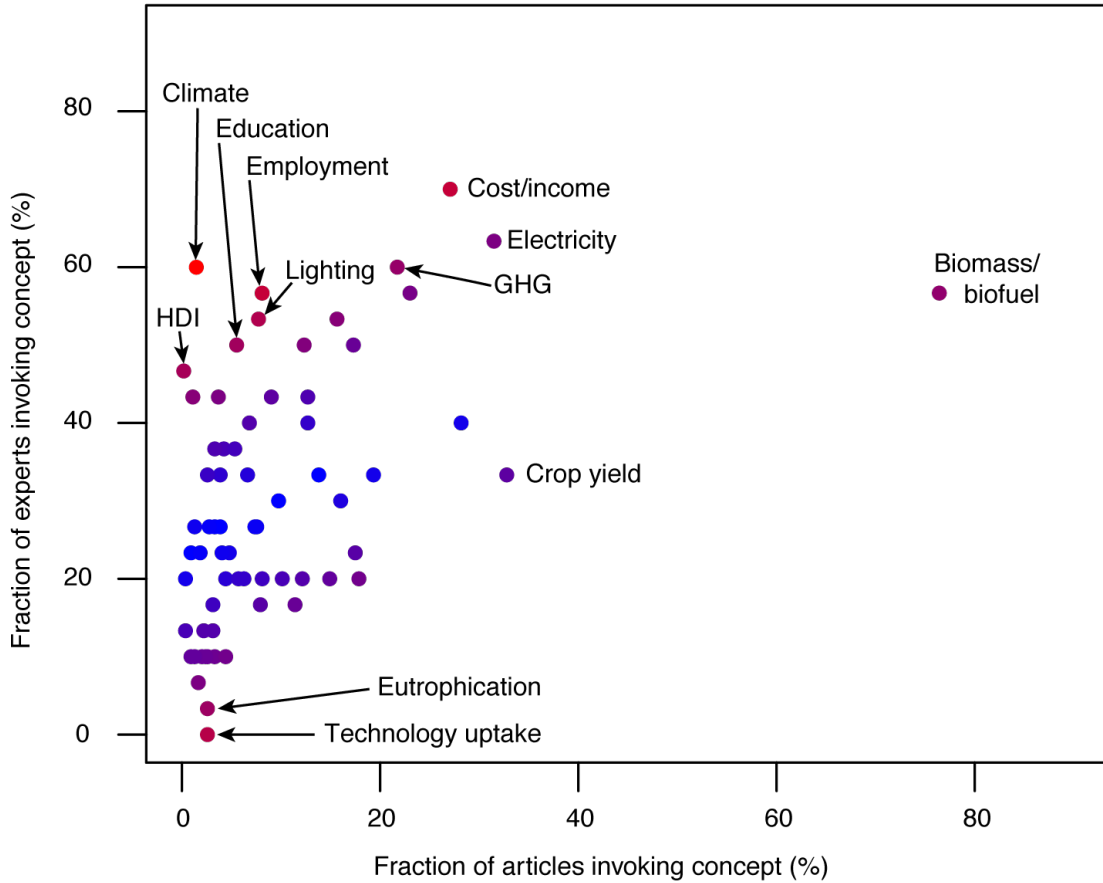


Figure S2: Concepts supplied to experts in elicitation workshops plotted according to the fraction of experts and the fraction of reviewed articles that invoked them. Color illustrates departure from linear regression (red = larger departure). Largest departures are labelled. Because concepts supplied to experts in the guided exercise accounted for a small subset of all concepts revealed in the literature review, most concepts had a higher percentage of experts than articles invoking them (average across concepts: 30% of experts vs. 9% of articles)



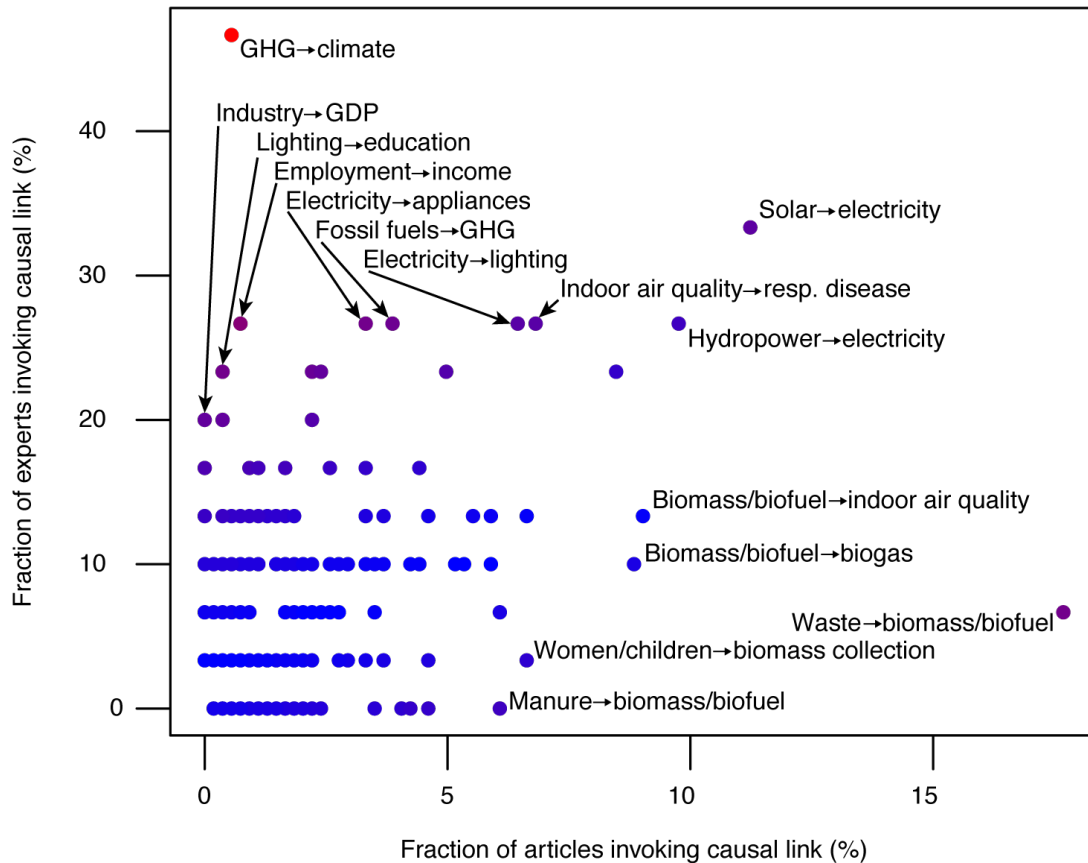


Figure S3: Relations drawn by experts from predefined concepts plotted according to the fraction of experts and the fraction of indexed articles that invoked them. Color illustrates departure from linear regression (red = larger departure). Largest departures are labelled. Because concepts supplied to experts in the guided exercise accounted for a small subset of all concepts revealed in the literature review, most links had a higher percentage of experts than articles invoking them (average across links: 4% of experts vs. 0.4% of articles)

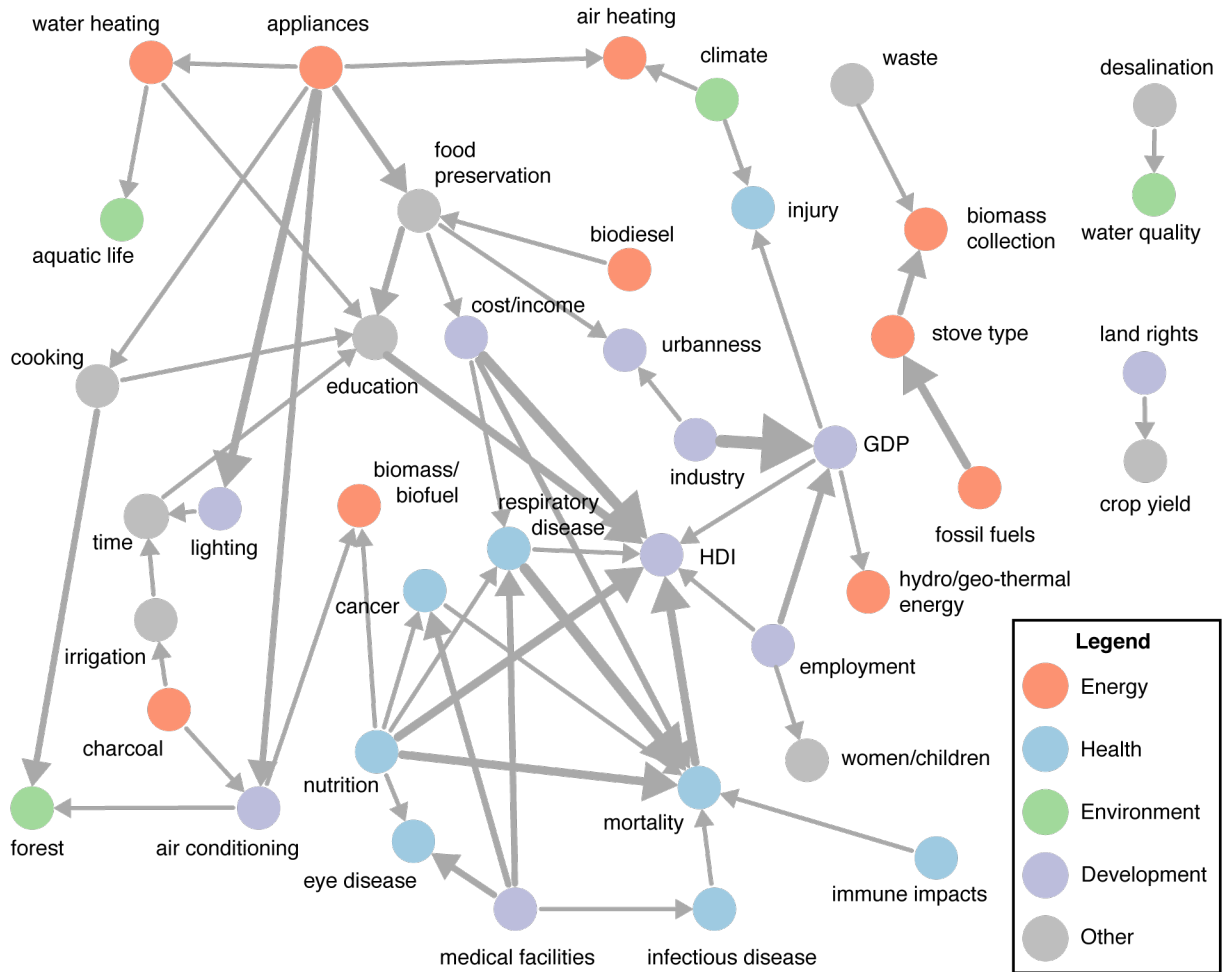


Figure S4: Relational network derived from guided expert elicitation exercises excluding connections already indexed in the literature (including via intermediate node) and connections identified by only one expert. Arrow width is proportional to number of experts identifying connection (range: 2–6).

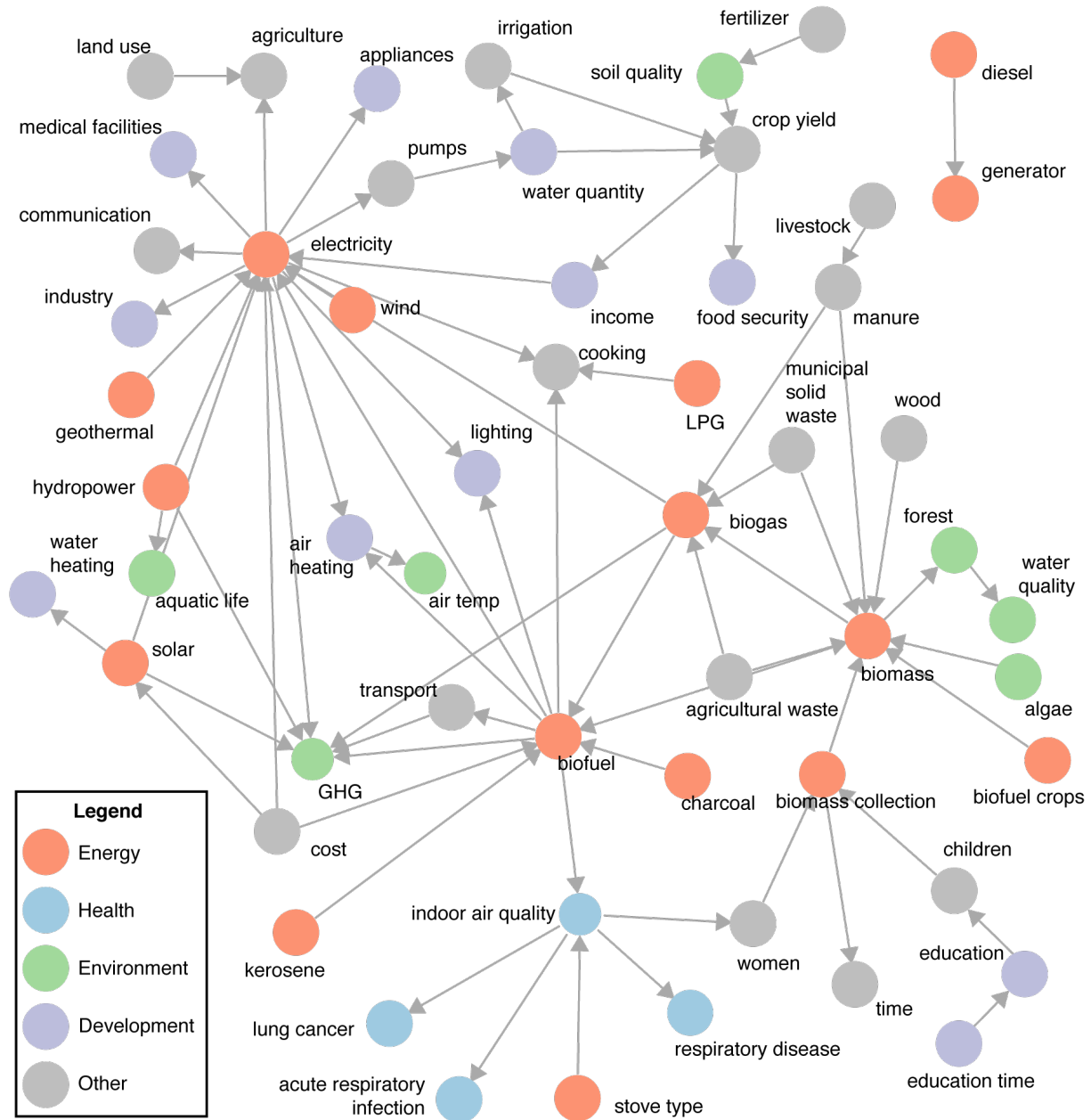


Figure S5: Relations mentioned in more than 10 articles (top 5% of occurrence). Nodes involved in relations mentioned in 10 or fewer articles are suppressed.

Table S1: Summary of evidence characterization

Association identified [sign of effect]	Review articles	Evidence type <sup>a</sup>					Primary refs <sup>b</sup>	Fraction of refs with no evidence	Fraction of refs with conflicting evidence <sup>c</sup>
		E	L	C	M	T			
agriculture → soil quality [-]	9	29	0	1	3	2	55	0%	22%
biodiesel → transport [+]	7	3	3	2	0	1	5	57%	0%
biofuel → electricity [+]	13	0	25	2	4	0	33	15%	0%
biofuel → GHG [+]	14	1	14	7	1	6	34	14%	29%
biofuel → indoor air quality [-]	40	1	14	2	5	1	67	18%	0%
biomass → forest [-]	32	4	20	1	2	4	82	13%	25%
biomass collection → time [-]	15	0	9	5	1	8	22	20%	0%
connectivity → biodiversity [+]	14	8	4	0	0	6	35	0%	7%
crop yield → income [+]	13	0	9	4	4	3	15	23%	8%
education time → education [+]	9	6	4	7	2	6	15	33%	0%
electricity → agriculture [+]	11	2	8	1	13	0	11	18%	9%
electricity → appliances [+]	15	1	3	1	9	8	27	27%	13%
electricity → communication [+]	12	2	6	0	13	0	18	33%	17%
electricity → cooking [+]	17	0	0	8	5	3	15	18%	41%
electricity → education [+]	7	0	6	4	2	1	7	57%	0%
electricity → employment [+]	10	0	5	0	9	1	18	20%	40%
electricity → income [+]	5	9	2	4	5	6	11	20%	38%
electricity → industry [+]	8	6	5	6	11	2	15	25%	38%
electricity → lighting [+]	31	1	1	9	1	0	42	16%	7%
electricity → medical facilities [+]	15	2	5	1	2	0	32	20%	7%
electricity → pumps [+]	20	0	3	8	0	2	32	40%	0%
electricity → water heating [+]	7	2	1	3	9	1	12	14%	14%
forest → biodiversity [+]	13	7	1	1	5	2	27	15%	0%
hydropower → biodiversity [-]	18	8	3	1	4	5	71	11%	11%
income → electricity [+]	13	1	8	4	1	0	26	0%	23%
lighting → education [+]	2	9	0	0	8	0	7	0%	0%
soil quality → biodiversity [+]	4	6	0	2	0	7	5	25%	25%
water quantity → irrigation [+]	16	0	0	3	4	2	25	13%	6%
women → biomass collection [+]	21	0	0	3	0	12	23	24%	0%

<sup>a</sup> C = correlational; E = experimental; L = longitudinal; M = modeling results; T = theories/inference; number corresponds to individual underlying references (primary sources) providing each type of evidence.

<sup>b</sup> Number of underlying references with evidence cited by the review articles evaluated in this study.

<sup>c</sup> Fraction of underlying references presenting evidence that disagrees with prevailing sign of effect.

## Appendix A: Web of Science search syntax

### Contents

1.	Search syntax for primary search.....	18
2.	Search syntaxes for secondary searches .....	19
2.1.	Air quality → biodiversity .....	19
2.2.	Eutrophication → nutrition .....	19
2.3.	Food waste → nutrition.....	19
2.4.	Forest → biodiversity.....	19
2.5.	Irrigation → crop yield.....	20
2.6.	Lighting or electricity → education.....	20
2.7.	Lighting or electricity → employment.....	20
2.8.	River fragmentation → biodiversity .....	21
2.9.	Soil quality → biodiversity .....	21
2.10.	Soil quality → crop yield .....	22
2.11.	Temperature control → nutrition.....	22
2.12.	Water quality → biodiversity .....	22

## 1. Search syntax for primary search

TS=(“energy transition\*” OR “sustainable energy” OR “household energy” OR “household fuel\*” OR "biomass fuel\*" OR "solid fuel\*" OR biogas OR LPG OR “liquefied petroleum gas” OR “transportation fuel\*” OR “clean\* fuel\*” OR “rural electrification” OR "electri\* access" OR firewood OR fuelwood OR charcoal OR photovoltaic\* OR solar\* OR “wind power\*” OR “alternat\* power” OR “alternat\* energy” OR "energy access" OR "access to energy" OR "energy security" OR microgrid\* OR microgrid\* OR “energy investment\*” OR “off grid” OR offgrid OR "hydro\* power" OR hydropower OR “micro hydro”)

AND

(

TS=(“public health” OR “human health” OR mortality OR morbidity OR disease OR illness OR incidence OR prevalence OR \*infect OR death OR injury)

OR

TS=(“economic development” OR “development goal\$” OR MDG OR poverty OR “well-being” OR livelihood\* OR trade OR wealth OR income)

OR

TS = (environment\* OR sustainab\* OR biodiversity OR ecosystem OR ecolog\* OR climate OR natur\* OR habitat)

)

AND

TS=(Africa\* OR "Latin America\*" OR "Central America\*" OR "South America\*" OR Asia\* OR "South Asia\*" OR "low income" OR "less developed countr\*" OR "developing countr\*" OR "emerging market\*" OR "emerging econom\*" OR "least developed countr\*" OR LIC\* OR LDC\* OR "third world" OR "global south" OR "middle income" OR LMIC OR "Afghanistan OR Albania OR Algeria OR Angola OR Argentina OR Armenia OR Azerbaijan OR Bangladesh OR Belarus OR Belize OR Benin OR Bhutan OR Bolivia OR Bosnia OR Botswana OR Brazil OR Bulgaria OR "Burkina Faso" OR Burundi OR "Cabo Verde" OR "Cape Verde" OR Cambodia OR Cameroon OR "Central African Republic" OR Chad OR China OR Colombia OR Comoros OR Congo OR "Costa Rica" OR "Côte d'Ivoire" OR “Cote d’Ivoire” OR “Ivory Coast” OR Cuba OR Djibouti OR Dominica OR "Dominican Republic" OR Ecuador OR Egypt OR "El Salvador" OR "Equatorial Guinea" OR Eritrea OR Ethiopia OR Fiji OR Gabon OR Gambia OR Georgia OR Ghana OR Grenada OR Guatemala OR Guinea OR Guinea-Bissau OR Guyana OR Haiti OR Honduras OR India OR Indonesia OR Iran OR Iraq OR Jamaica OR Jordan OR Kazakhstan OR Kenya OR Kiribati OR "North Korea" OR DPRK OR Kosovo OR Kyrgyz\* OR Lao\* OR Lebanon OR Lesotho OR Liberia OR Libya OR Macedonia OR Madagascar OR Malawi OR Malaysia OR Maldives OR Mali OR "Marshall Islands" OR Mauritania OR Mauritius OR Mexico OR Micronesia OR Moldova OR Mongolia OR Montenegro OR Morocco OR Mozambique OR Myanmar OR Namibia OR Nepal OR Nicaragua OR Niger OR Nigeria OR Pakistan OR Palau OR Panama OR "Papua New Guinea" OR Paraguay OR Peru OR Philippines OR Romania OR Russia\* OR Rwanda OR Samoa OR “São Tomé and Príncipe” OR “São Tomé and Príncipe” OR “Sao Tome and Principe” OR Senegal OR Serbia OR "Sierra Leone" OR

"Solomon Islands" OR Somalia OR "South Africa" OR "South Sudan" OR "Sri Lanka" OR "St\* Lucia" OR "Saint Lucia" OR "St\* Vincent" OR "Saint Vincent" OR "Saint Vincent and the Grenadines" OR "St\* Vincent and the Grenadines" OR Sudan OR Suriname OR Swaziland OR Syria OR Tajikistan OR Tanzania OR Thailand OR Timor-Leste OR "East Timor" OR Togo OR Tonga OR Tunisia OR Turkey OR Turkmenistan OR Tuvalu OR Uganda OR Ukraine OR Uzbekistan OR Vanuatu OR Venezuela OR Vietnam OR "West Bank" OR Gaza\* OR Yemen OR Zambia OR Zimbabwe)

AND

SU= (ENERGY FUELS OR SCIENCE TECHNOLOGY OTHER TOPICS OR ENVIRONMENTAL SCIENCES ECOLOGY OR ENGINEERING OR AGRICULTURE OR PUBLIC ENVIRONMENTAL OCCUPATIONAL HEALTH OR SOCIAL SCIENCES OTHER TOPICS OR SOCIAL ISSUES)

Refined by: DOCUMENT TYPES: (REVIEW)

## 2. Search syntaxes for secondary searches

### 2.1. Air quality → biodiversity

**Results:** 22 (from Web of Science Core Collection)

**You searched for:** TS=("biodiversity") AND TS="air quality"

**Refined by:** DOCUMENT TYPES: ( REVIEW )

**Timespan:** All years. **Indexes:** SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC.

### 2.2. Eutrophication → nutrition

**Results:** 24 (from Web of Science Core Collection)

**You searched for:** TS=("eutrophication") AND TS=("Nutrition")

**Refined by:** DOCUMENT TYPES: ( REVIEW )

**Timespan:** All years.

**Indexes:** SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC.

### 2.3. Food waste → nutrition

**Results:** 11 (from Web of Science Core Collection)

**You searched for:** TS=("Food waste" OR "food spoilage") AND TS=("Nutrition")

**Refined by:** DOCUMENT TYPES: ( REVIEW )

**Timespan:** All years.

**Indexes:** SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC.

### 2.4. Forest → biodiversity

**Results:** 26 (from Web of Science Core Collection)

**You searched for:** TS=("biodiversity") AND TS=("forest cover" OR "forest density")

**Refined by:** DOCUMENT TYPES: ( REVIEW )

**Timespan:** All years.

**Indexes:** SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC.

2.5. Irrigation → crop yield

**Results:** 48 (from Web of Science Core Collection)

**You searched for:** TS=("crop yield") AND TS=("irrigation" OR "pumps")

**Refined by:** DOCUMENT TYPES: ( REVIEW )

**Timespan:** All years.

**Indexes:** SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC.

2.6. Lighting or electricity → education

**Results:** 75 (from Web of Science Core Collection)

**You searched for:** TS=("electricity" OR "lighting") AND TS=("education")

**Refined by:** DOCUMENT TYPES: ( REVIEW )

**Timespan:** All years.

**Indexes:** SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC.

2.7. Lighting or electricity → employment

**Results:** 169 (from Web of Science Core Collection)

**You searched for:**

TS=("electricity" OR "lighting")

AND

TS=("entrepreneurship" OR "business" OR "entrepreneur" OR "employment" OR "job" OR "jobs" OR "wage labor" OR "wage labour" OR income)

AND

TS=(Africa\* OR "Latin America\*" OR "Central America\*" OR "South America\*" OR Asia\* OR "South Asia\*" OR "low income" OR "less developed countr\*" OR "developing countr\*" OR "emerging market\*" OR "emerging econom\*" OR "least developed countr\*" OR LIC\* OR LDC\* OR "third world" OR "global south" OR "middle income" OR LMIC OR "Afghanistan OR Albania OR Algeria OR Angola OR Argentina OR Armenia OR Azerbaijan OR Bangladesh OR Belarus OR Belize OR Benin OR Bhutan OR Bolivia OR Bosnia OR Botswana OR Brazil OR Bulgaria OR "Burkina Faso" OR Burundi OR "Cabo Verde" OR "Cape Verde" OR Cambodia OR Cameroon OR "Central African Republic" OR Chad OR China OR Colombia OR Comoros OR Congo OR "Costa Rica" OR "Côte d'Ivoire" OR "Cote d'Ivoire" OR "Ivory Coast")



OR Cuba OR Djibouti OR Dominica OR "Dominican Republic" OR Ecuador OR Egypt OR "El Salvador" OR "Equatorial Guinea" OR Eritrea OR Ethiopia OR Fiji OR Gabon OR Gambia OR Georgia OR Ghana OR Grenada OR Guatemala OR Guinea OR Guinea-Bissau OR Guyana OR Haiti OR Honduras OR India OR Indonesia OR Iran OR Iraq OR Jamaica OR Jordan OR Kazakhstan OR Kenya OR Kiribati OR "North Korea" OR DPRK OR Kosovo OR Kyrgyz\* OR Lao\* OR Lebanon OR Lesotho OR Liberia OR Libya OR Macedonia OR Madagascar OR Malawi OR Malaysia OR Maldives OR Mali OR "Marshall Islands" OR Mauritania OR Mauritius OR Mexico OR Micronesia OR Moldova OR Mongolia OR Montenegro OR Morocco OR Mozambique OR Myanmar OR Namibia OR Nepal OR Nicaragua OR Niger OR Nigeria OR Pakistan OR Palau OR Panama OR "Papua New Guinea" OR Paraguay OR Peru OR Philippines OR Romania OR Russia\* OR Rwanda OR Samoa OR "São Tomé and Príncipe" OR "São Tomé and Príncipe" OR "Sao Tome and Principe" OR Senegal OR Serbia OR "Sierra Leone" OR "Solomon Islands" OR Somalia OR "South Africa" OR "South Sudan" OR "Sri Lanka" OR "St\*Lucia" OR "Saint Lucia" OR "St\* Vincent" OR "Saint Vincent" OR "Saint Vincent and the Grenadines" OR "St\* Vincent and the Grenadines" OR Sudan OR Suriname OR Swaziland OR Syria OR Tajikistan OR Tanzania OR Thailand OR Timor-Leste OR "East Timor" OR Togo OR Tonga OR Tunisia OR Turkey OR Turkmenistan OR Tuvalu OR Uganda OR Ukraine OR Uzbekistan OR Vanuatu OR Venezuela OR Vietnam OR "West Bank" OR Gaza\* OR Yemen OR Zambia OR Zimbabwe)

**Refined by:** DOCUMENT TYPES: ( REVIEW )

**Timespan:** All years.

**Indexes:** SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC.

#### 2.8. River fragmentation → biodiversity

**Results:** 2 (from Web of Science Core Collection)

**You searched for:** TS=("biodiversity") AND TS=("river fragmentation" OR "river segmentation")

**Refined by:** DOCUMENT TYPES: ( REVIEW )

**Timespan:** All years.

**Indexes:** SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC.

#### 2.9. Soil quality → biodiversity

**Results:** 50 (from Web of Science Core Collection)

**You searched for:** TS=("soil quality") AND TS="biodiversity"

**Refined by:** DOCUMENT TYPES: ( REVIEW )

**Timespan:** All years.

**Indexes:** SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC.

2.10. Soil quality → crop yield

**Results:** 33 (from Web of Science Core Collection)

**You searched for:** TS=("soil quality") AND TS="crop yield"

**Refined by:** DOCUMENT TYPES: ( REVIEW )

**Timespan:** All years.

**Indexes:** SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC.

2.11. Temperature control → nutrition

**Results:** 48 (from Web of Science Core Collection)

**You searched for:** TS=("refrigeration" OR "heating" OR "cooling" OR "air conditioning" OR "HVAC") AND TS=("food security" OR "nutrition")

**Refined by:** DOCUMENT TYPES: ( REVIEW )

**Timespan:** All years.

**Indexes:** SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC.

2.12. Water quality → biodiversity

**Results:** 153 (from Web of Science Core Collection)

**You searched for:** TS=("biodiversity") AND TS="water quality"

**Refined by:** DOCUMENT TYPES: ( REVIEW )

**Timespan:** All years. **Indexes:** SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC.

## Appendix B: Consent materials for expert elicitation workshops

### Contents

Expert elicitation recruitment email.....	24
Expert elicitation follow-up email .....	25
Expert elicitation instructions script .....	26
1. Introduction and instructions.....	26
2. Interventions, effects and outcomes of interest (non-directive).....	28
3. Predefined list of interventions, effects and outcomes (directive) .....	28
Participant consent form.....	29
1. Why is this research study being done?.....	29
2. Who is conducting this research? .....	29
3. What will I be asked to do?.....	29
4. How long will I be in the study? .....	29
5. What are the risks, inconveniences and benefits of this study? .....	30
6. Confidentiality.....	30
7. Voluntary nature of participation .....	30
8. Interview recording.....	30
9. Whom do I call if I have questions or problems?.....	30
Duke University Institutional Review Board Approval.....	30

### Expert elicitation recruitment email

**From:** Ryan Calder <[ryan.calder@duke.edu](mailto:ryan.calder@duke.edu)>

**Subject:** SETI Conference: we need your perspectives

**Date:** May 14, 2018 at 7:25 AM

**To:** «First» «Last» <<Email>>

Dear «First» «Last»:

We are looking forward to seeing you at the Sustainable Energy Transitions Initiative (SETI) at Duke University next week. Given your expertise, **we would like to invite you to participate in a 30-to-45-minute interview** during your time at Duke, scheduled at your convenience.

As you know, increasing access to energy is an important part of overall plans to boost economic and social development around the world. The interconnected nature of food, energy and water systems means that interventions in one area may have unexpected effects in another area. We are conducting a research project to characterize these interacting effects using the published literature, available data and verbal descriptions from expert practitioners like you. Our research team consists of faculty and staff from across Duke who are devoted to interdisciplinary approaches for understanding and addressing global development, environmental and health challenges.

We would like to ask you what interventions you work with, what effects they have, and how these and other effects combine to impact development, environment and health outcomes. We will use the information we collect from experts like you to identify possible gaps in (and differences with) the published literature to scope future research questions and proposals for development efforts.

**Your participation is strictly voluntary** and will be kept confidential. We will record and report only anonymized data.

We hope you will choose to participate. Please reply to study coordinator Ryan Calder ([ryan.calder@duke.edu](mailto:ryan.calder@duke.edu)) to schedule an interview time, or with any questions on this research.

On behalf of the research team, I thank you for your consideration and look forward to seeing you at Duke.

- Marc Jeuland, Associate Professor of Public Policy and Global Health
- Mark Borsuk, Associate Professor of Civil and Environmental Engineering
- Kyle Bradbury, Lecturing Fellow and Managing Director, Energy Data Analytics Lab, Duke Energy Initiative
- Jordan Malof, Assistant Research Professor of Electrical and Computer Engineering
- Lydia Olander, Ecosystem Services Program Director, Nicholas Institute for Environmental Policy Solutions
- Ryan Calder, Postdoctoral Associate, Civil and Environmental Engineering
- Rob Fetter, Senior Policy Associate, Nicholas Institute for Environmental Policy Solutions
- Jonathan Phillips, Energy Access Program Director, Nicholas Institute for Environmental Policy Solutions

**Expert elicitation follow-up email**

**From:** Ryan Calder <[ryan.calder@duke.edu](mailto:ryan.calder@duke.edu)>

**Subject: Re: SETI Conference: we need your perspectives**

**Date:** May 16, 2018 at 11:50 AM

**To:** «First» «Last» <<Email>>

Dear «First» «Last»:

On Monday, we sent you an invitation to participate in a **30-to-45-minute** research interview during the Sustainable Energy Transitions Initiative (SETI) conference at Duke.

We would like to use this interview to characterize how you, as an expert practitioner, think about interacting challenges in international development.

**Your participation is strictly voluntary.** If you choose to participate, or if you have any questions about this research project, please reply to study coordinator Ryan Calder ([ryan.calder@duke.edu](mailto:ryan.calder@duke.edu)).

Looking forward to meeting you.

**Expert elicitation instructions script**  
Rev. 2018-05-02

**1. Introduction and instructions**

“The purpose of this interview is to understand how you, as an expert practitioner, think about cause and effect relationships between energy access and societal outcomes in your practice area. This means we want to understand what actions lead to what effects, what other factors may influence those effects, and how the various effects ultimately lead to societal outcomes of interest. We will be doing this using a type of drawing called an influence diagram.

“As an everyday example, an influence diagram can be drawn to show how disciplining children can reduce the risk of falling down the stairs by reducing the likelihood of their leaving toys on the stairs and thus reducing the risk of tripping. This diagram also shows that there are a number of other factors that change the risk of tripping (and falling) on the stairs. Some of these factors interact with each other. For example, keeping the stairs well-lit may directly reduce the risk of tripping and also decrease the likelihood that the cat will sleep on the stairs, reducing the risk of tripping on the cat. If you are more agile, you are less likely to trip and also less likely to fall in the event that you do trip. There can also be feedbacks. For example, frequent tripping on the stairs may make children more likely to clean up their toys.

“In this exercise, I am going to ask you to enumerate the actions, effects and outcomes that you study and then ask you to arrange them graphically according to the influences they exert on each other. Then, I am going to provide you with a predefined list of these concepts and ask you to arrange them according to causal relations.

“Before we start, I have to record your consent to participate in this study. This consent form is the only place your name will appear, and it will not be published or even digitized or stored with the survey results.

“Do you have any questions before we proceed?”