Studying the effects of intervention programmes on household energy saving behaviours using graphical causal models

Nitin Bhushan, Linda Steg, and Casper Albers
Heymans Institute for Psychological Research
Faculty of Behavioural and Social Sciences
University of Groningen
Grote Kruisstraat 2/1
9712TS Groningen
The Netherlands
Corresponding author e-mail: n.bhushan@rug.nl

Randomised controlled trials are strongly advocated to evaluate the effects of intervention programmes on household energy saving behaviours. While randomised controlled trials are the ideal, in many cases, they are not feasible. Notably, many intervention studies rely on voluntary participation of households in the intervention programme, in which case random selection and random assignment are seriously challenged. Moreover, studies employing randomised controlled trials typically do not study the underlying processes causing behaviour change. Yet, the latter is highly important to improve theory and practice. We propose a systematic approach to causal inference based on graphical causal models to study effects of intervention programmes on household energy saving behaviours when randomised controlled trials are not feasible. Using a simple example, we explain why such an approach not only provides a formal tool to accurately establish effects of intervention programmes, but also enables a better understanding of the processes underlying behaviour change.

Keywords: Causality, intervention evaluation, graphical causal models, directed acyclic graphs, confounding, collider bias

To mitigate anthropogenic climate change, households across the world need to reduce their fossil energy consumption and engage in energy saving behaviours (IPCC, 2014). To this end, reviews and meta-analyses show that various behavioural intervention programmes including block leader approaches, behavioural commitments, and different types of feedback appear to encourage household energy saving behaviours (Abrahamse, Steg, Vlek, & Rothengatter, 2005; Karlin, Zinger, & Ford, 2015). Typically, studies that aimed to examine the effects of such interventions did not follow rigorous study designs, and did not try to understand the processes that lead to the observed effects, so little is known about why intervention programmes are (in)effective and how they can be improved (Abrahamse et al., 2005). Considerable improvements are possible in the design of intervention programmes to not only evaluate, but also understand the effects of such interventions on household energy saving behaviours.

One way of ensuring that any change in energy usage can be attributed to the intervention programme and nothing else is by conducting a Randomised Controlled Trial (RCTs), also termed as true experiments. Recently, RCTs have been strongly advocated to evaluate intervention programmes in the household energy efficiency domain (Allcott & Mullainathan, 2010; Frederiks, Stenner, Hobman, & Fischle, 2016; Vine, Sullivan, Lutzenhiser, Blumstein, & Miller, 2014). RCTs allow drawing firm conclusions about the extent to which intervention programmes are effective in encouraging households to realise energy savings because of three key characteristics: (i) manipulation; (ii) random sampling of households from the target population; and (iii) random assignment of households to intervention and control groups. Manipulation implies some households are deliberately exposed to the intervention while a control group does not receive the intervention. Control groups are essential to test whether any changes in energy use can be attributed to the intervention, and not to any other event happening during the test of the intervention. Random assignment ensures that the intervention and control groups do not systematically differ from the outset, and ensure that changes in energy use are not caused by specific characteristics of the intervention group. Furthermore, random sampling ensures that results can be generalised to the target population.

The proponents of RCTs argue that if the three features are rigorously implemented, RCT's enable accurate evaluation of the effects of an intervention programme on energy saving behaviours. In simpler terms, when researchers and policy makers are interested in finding out *"if"* the intervention programme worked, RCTs provide the best answer (Lilienfeld, McKay, & Hollon, 2018).

However, in the context of household energy use intervention programmes, various real-world constraints do not permit use of RCTs (Vine et al., 2014). These real-world constraints imply that certain methodological challenges arising due to the infeasibility of conducting RCTs may not just be inadvertent, but also unavoidable. For example, when one would like to study effects of doubling of energy costs on energy usage, regulatory, institutional, and ethical constraints may not allow random assignment of participants to intervention and control groups. Moreover, due to legal and privacy constraints, most intervention programmes imply that people have to sign up and agree to participate, which challenges random sampling and random assignment. Hence, key elements of RCTs – random selection and random assignment – are often not feasible in real life. This implies one can no longer rule out the possibility that participants in the study are not a representative sample of the target population, or that intervention and control groups do not systematically differ from the outset. This may result in inaccurate estimates of the effects of the intervention programme on household energy saving behaviours, as it is not clear whether results can be generalized to the target population, or whether any differences in energy behaviour after the interventions are caused by the intervention programme, and not by other systematic differences between intervention and control groups.

Such real-world constraints imply that conducting RCTs is not always feasible in practice. In addition, most studies employing RCTs estimate the effects of intervention programmes without trying to understand the processes that underlie the effects of such interventions. As such, one of the key drawback of RCTs is that they do not improve our understanding of *"why"* these programmes work (Carey & Stiles, 2016; Deaton & Cartwright, 2016; Vandenbroucke, 2008). Understanding the processes through which intervention programmes affect energy saving behaviours is important to improve the design of such programmes and to advance scientific theory. For example, tailored information campaigns to promote energy saving behaviours may be effective because they enhance knowledge about energy saving options, or maybe because information that aligns with what people find important strengthens one's motivation to save energy. To study processes underlying intervention effects, one would need to collect information on relevant process variables (e.g. knowledge, motivation), which in many cases have to be collected via questionnaires. Here, one again has to rely on voluntary participation of participants, challenging random sampling and random assignment, and making RCTs infeasible.

Hence, real-life circumstances often challenge the feasibility of RCTs. Yet, this does not imply that researchers cannot carefully evaluate the effects of an intervention programme on household energy saving behaviours (Shadish, Clark, & Steiner, 2008). When randomisation is not feasible, there are several empirical alternatives to RCTs (for reviews of alternatives, see Carey and Stiles 2016; Cook, Campbell, and Shadish 2002; Vine et al. 2014; West 2009). One such alternative to RCTs, living labs, implies that causal inference is challenging, as typically, no random assignment or random selection takes place. Another commonly adopted alternative, quasi-experiments is used when random assignment is not feasible. A key drawback of such designs is that the lack of random assignment implies that we cannot rule out alternative explanations for the observed intervention effect, which leads to bias in evaluating the effects of the intervention programme on energy saving behaviours. Typically, researchers aim to rule out these alternative explanations and minimize bias by controlling for third variables which are supposed to be related to both partaking in the intervention programme and energy saving behaviours. However, as we show in this paper, this does not always minimise bias in the evaluation of the effects of the intervention programme on household energy saving behaviours, and perhaps non-intuitively, controlling for such variables may even induce bias in evaluating the effects of the intervention programme on household energy saving behaviours.

Hence, careful examination of the effects of an intervention programme on household energy saving behaviours in the absence of randomisation requires a systematic approach to dealing with bias. Moreover, similar to RCTs, while non-experimental designs such as quasi-experiments might permit researchers to evaluate effects of intervention programmes on energy saving behaviours, they do not necessarily provide insights in why these interventions were effective or not, which is key to understanding and designing better interventions. Hence, an important question is: Which would be an appropriate second best solution to to carefully evaluate the effects of intervention programmes on household energy saving behaviours when RCTs are not feasible by systematically approaching bias, that also improves our understanding of the processes underlying the effects of the programme?

Graphical Causal Models

We propose that graphical causal models, and in particular, causal directed acyclic graphs (DAGs), offers a promising second-best approach to evaluate and understand effects of intervention programmes on household energy saving behaviours when RCTs are not feasible (Pearl, 2009; Spirtes, Glymour, & Scheines, 2000). A DAG consists of a set of variables (so-called nodes) and a set of lines (so-called edges) denoting relationships between the variables. In a DAG, the directed edges (i.e. one directional arrows) represent causal paths between variables. For example, a directed line from partaking in an intervention programme to household energy saving behaviours implies that the intervention programme has a direct causal effect on household energy saving behaviours.

In household energy studies, a DAG is an explicit description of the causal mechanisms underlying effects of intervention programmes on household energy saving behaviours and is based on scientific theory. In a way, DAGs are similar to path models that are more widely used to study household energy saving behaviours, but there are some differences between the two. Notably, DAGs encode qualitative assumptions about how the intervention affects behaviours, and a directed line between two variables in a DAG represents the causal effect between the variables irrespective of the type of the effect (e.g. linear, quadratic, cubic). Hence path models, which generally model linear causal effects, can be classified as a specific instance of a DAG.

A key advantage of using DAGs is that it forces researchers to systematically consider possible biases that may obscure the true effect of an intervention programme on household energy saving behaviours (Greenland, Pearl, & Robins, 1999; Shrier & Platt, 2008). In the absence of random assignment, as is often the case in quasi-experimental designs, the traditional approach to minimize bias in evaluating the effect of the intervention programme on household energy saving behaviours is to statistically control for all variables which could influence energy saving behaviours next to the intervention by including the variables as co-variates in a regression or path model. However, statistical controlling (henceforth, controlling) for related variables does not always minimize bias in the effect of the intervention programme on household energy saving behaviours and perhaps non-intuitively, controlling for such variables may even induce bias in evaluating the effect of the intervention programme on household energy saving behaviours.

When randomization is not feasible, two major types of biases can affect the accuracy of evaluating the effect of the intervention programme on household energy saving behaviours: confounding biases and collider biases. Confounding biases are due to factors that influence participation in the intervention programme as well as household energy behaviours. On the contrary, collider biases are due to factors influenced by participation in the intervention programme as well as household energy behaviours (see Table 1 for a summary of key terms used in this paper).

We illustrate these two biases using DAGs. Figure 1(a) is a DAG based on theory that represents

Table 1
Definition of key terms

Term	*Description*
RCT	Randomised Controlled Trial. Involves manipulation, random selection, and random assignment.
DAG	Directed Acyclic Graph. A systematic approach to evaluate effects of interventions when RCTs are not possible. A DAG consists of a set of variables (so-called nodes) and a set of lines (so-called edges) denoting relationships between the variables.
Confounder	A variable that affects partaking in an intervention programme (the independent variable) as well as energy saving behaviour (the dependent variable)
Collider	a variable that is affected by partaking in an intervention programme (the independent variable) as well as by energy saving behaviour (the dependent variable)

the processes underlying the effects of feedback on household energy saving behaviours. Environmental concern is theorized to cause participation in the feedback programme as well as engagement in energy saving behaviours. Furthermore, it was theorized that participation in the feedback programme strengthens motivation to save energy, and that increased engagement in energy saving behaviour also strengthens this motivation.

In this example, environmental concern is a confounder that tends to mask the real relationship between the feedback programme and household energy saving behaviours (denoted by the dotted line). As shown in Figure 1(b), statistically controlling for environmental concern by including the variable as a co-variate in a regression model

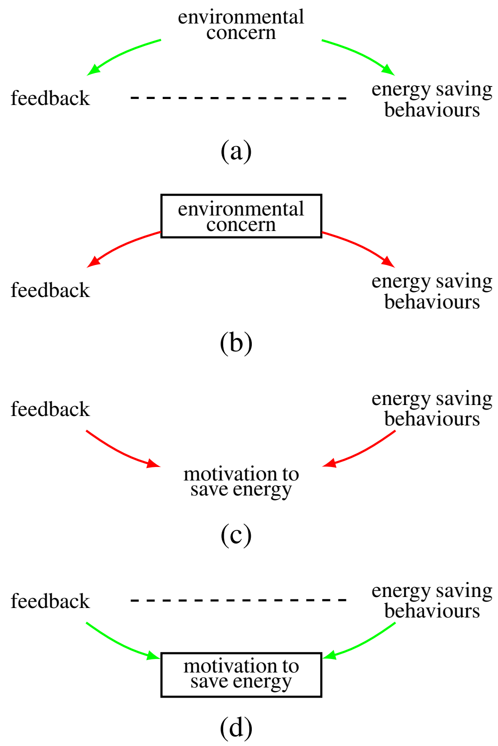


Figure 1. DAG illustrating bias due to confounding (1(a) and 1(b)) and a collider (1(c) and 1(d)). Statistically controlling for a confounder minimizes bias in estimating the effect of an intervention programme on energy saving behaviour. Statistically controlling for a collider can induce bias in estimating the effect of an intervention programme on household energy saving behaviours. Note: boxes around a variable denote statistical control and dotted lines represent spurious correlations

would block any spurious relationships between feedback and energy saving behaviours, and would thus minimise confounding bias while estimating the effect of the feedback programme on household energy saving behaviours.

Collider biases imply that feedback as well as engaging in household energy saving behaviours influence a third variable; controlling for this third variable would induce a spurious relation between feedback and energy saving behaviours as it would suggest that feedback has an effect on household energy use even when there is no true effect. In our example for a collider bias, we observe that motivation to save energy is caused by both feedback and engaging in energy saving behaviours. Let's assume that feedback has no effect on household engagement in energy saving behaviours. Now, controlling on the collider, motivation to save energy, is equivalent to looking at the effect of feedback on household energy saving behaviours only among highly motivated households. This leads to a spurious relation between feedback and energy saving behaviours and is termed as collider bias (for more examples of collider bias, see Cole et al. 2010; Elwert and Winship 2014).

These examples illustrate that controlling for a third variable in a model can sometimes change (i.e. remove, induce, or change the direction of) the association between any two other variables related to a third variable in the model. This is termed as Simpson's paradox and Berksons's paradox. More generally, the paradox states that the direction of an association at the population-level may be reversed when examined in subgroups within the population (Albers, 2015; Kievit, Frankenhuis, Waldorp, & Borsboom, 2013). Using DAGs on the basis of a clear theory describing how an intervention programme may affect energy saving behaviour will prevent such biases and paradox (Pearl, 2014).

Hence, a key question faced by researchers when evaluating the effect of an intervention programme on household energy saving behaviours when randomisation is not feasible is: What variables should we control for in order to minimize bias, and what variables should we not control for to inadvertently induce bias?

In the following section we show how DAGs can help answer this question (see Pearl (2009) for technical details of this method). Given a DAG, several software packages can be used to determine what variables to include in order to carefully evaluate effects of intervention programmes on household energy saving behaviours based on graphical causal models. Commonly used R (R Development Core Team, 2008) packages include pcalg (Kalisch, Mächler, Colombo, Maathuis, & Bühlmann, 2012) and dagitty (Textor, van der Zander, Gilthorpe, Liśkiewicz, & Ellison, 2016). In addition, as an alternative to the R package, a web application "DAGitty" is easy to use and freely available at `http://dagitty.net`.

We propose a systematic approach (see Figure 2) based on DAGs to carefully conduct and evaluate the effect of an intervention programme on household energy saving behaviours. We break down the process into four steps: (i) explicate a theoretical model that explains how the intervention programme affects household energy saving behaviours, (ii) draw a DAG representing the theoretical model and identify what factors must be controlled for in order to estimate the causal effect of the intervention programme, (iii) implement the programme and collect data on energy saving behaviours and all relevant process variables identified in the previous step, (iv) and estimate the effects of the intervention programme on household energy saving behaviours.

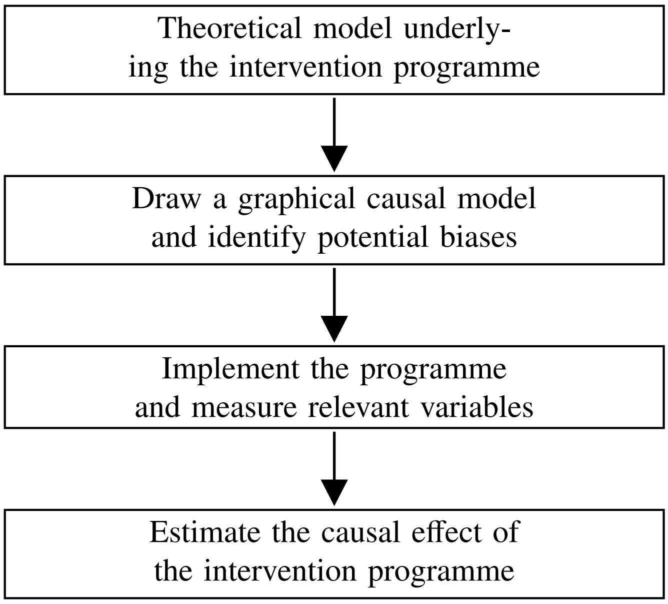


Figure 2. A systematic approach based on graphical causal models to design and evaluate effects of an intervention programme on household energy saving behaviours when RCTs are not feasible.

Example

In this section, we use a simple example to illustrate how one can use DAGs and simple web based software such as "DAGitty" to minimise bias in estimating causal effect of intervention programmes on household energy savings. We would like to emphasize that this is a simple example with the goal to introduce and illustrate how a systematic approach based on DAGs can help minimise bias in evaluating effects of intervention programmes on household energy saving behaviours. The example is intentionally kept simple to illustrate the concepts underlying causal inference with DAGs. For theories that involve a few more variables, the same mechanisms can still be applied. In case there are many variables (e.g., dozens), things do become more complicated (cf. Shrier and Platt 2008), but most theories describing how interventions affect energy saving behaviour are not concerned with dozens of variables at the same time. When many variables are involved, causal identification methods based on DAGs (e.g. backdoor algorithm) can be use to perform this very task accurately (Pearl, 2009).

Consider an intervention programme that aims to examine to what extent providing households with information on the negative environmental impact of their energy use (the intervention) will encourage them to engage in energy saving behaviours. Randomisation is not feasible as households can choose whether to sign up and be part of this programme.

Step 1: Theoretical model underlying the effects of the intervention programme

First, based on theory, we assume that partaking in the intervention programme will result in household energy saving behaviour by increasing participants' awareness of the environmental impact of their energy use behaviours. This implies that participants' awareness of the environmental impact of their energy use is expected to mediate the effect of the intervention on their energy saving behaviours. In addition, we theorize that households are more likely to participate in the intervention programme when they care more about the environment. Furthermore, people are more likely to engage in energy saving behaviours when they care more about the environment. Here, environmental concern affects the likelihood of participation in the programme as well as the likelihood of engaging in energy saving behaviours and is therefore a confounder. Hence, in order to minimise bias in estimating the effect of the intervention programme on energy saving behaviours, we must control for environmental concern.

Furthermore, we theorize that knowledge about effective ways to reduce energy savings may be increased due to participating in the programme, as participants may look for energy saving tips. Yet, such knowledge may also result from engagement in energy saving behaviours, when people notice reductions in energy use because of changes in specific behaviours. This implies that increase in knowledge of effective ways to reduce energy may be caused by participation in the programme, but also by energy savings realised due to engagement in energy saving behaviours. Knowledge is thus a collider, and we must not control for knowledge in order to accurately estimate the effect of partaking in the intervention programme on household energy saving behaviours.

Step 2: Draw a graphical causal model and identify potential biases

Next, we draw a DAG based on our theoretical reasoning underlying the effects of the intervention programme. We use DAGitty to draw the DAG and Figure 3 displays the resulting DAG. DAGitty can also be used to identify any potential lurking sources of biases and is useful when researchers draw complex causal models with variables originating from multiple theories. In our simple example, as we earlier identified, DAGitty indicates that (given this DAG) environmental concern must be controlled for in order to accurately estimate the effect of partaking in the intervention programme

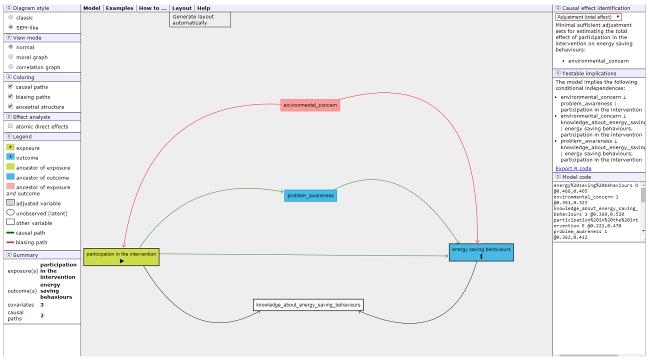


Figure 3. Screenshot of the results obtained from DAGitty. Note that the tab displaying causal effect identification indicates what variables must be controlled for in order to carefully estimate causal effects of the intervention on energy saving behaviours. Image source: http://dagitty.net

on energy saving behaviours (as it is a confounder) and knowledge must not be controlled for (as it is a collider).

Step 3: Implement the intervention programme and measure relevant variables

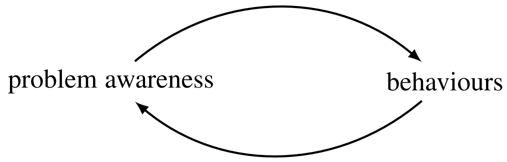
Now that the theoretical model has been specified, and relevant confounders and colliders have been identified, we can now implement the intervention programme and collect data on the model variables and energy saving behaviours. Assume that 200 households chose to participate in the intervention programme (response rate of 30%); and provide access to their electricity meter readings. In addition, they also complete a questionnaire a week before the start of the programme, and five months after the start of the intervention intervention measuring their level of environmental concern, and problem awareness. Note that we do not measure knowledge of energy saving behaviours as this is a collider.

Step 4: Estimate the effect of the intervention programme on energy saving behaviours

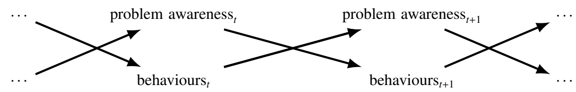
In the final step, to estimate the causal effect of the intervention on energy saving behaviours, a path model is specified with household energy saving behaviours as the dependent variable, partaking in the intervention as the independent variable, and problem awareness as the mediator; and to minimise bias, we control for environmental concern in the analysis. After fitting the model, the path coefficient of partaking in the intervention programme can be interpreted as the total causal effect of participation on energy saving behaviours.

Dynamic graphical causal models

So far, we have described graphical causal models which can help estimate static causal effects. However, intervention effects may often change with time. Hence, it may be important to study the effects of intervention programmes on household energy savings as a dynamic process, in which changes in energy saving behaviours (short-term and long-term effects) as well as changes in underlying determinants of the behaviour over time are systematically evaluated. Using longitudinal measurements, dynamic graphical causal models enable to assess how changes in behavioural antecedents affect changes in household energy saving behaviours and hence, long term effects of interventions can be examined using these models (Greenland et al., 1999). Another limitation of DAGs is that they are acyclic and do not allow for feedback loops that may affect household energy saving behaviours. Feedback and reciprocal causation can also be represented using dynamic graphical causal models. When time is explicitly taken into account (e.g. by longitudinal measurements), DAGs can model feedback processes. see Figure 4 for an example of dynamic graphical causal models. Figure 4(b) is a dynamic representation of Figure 4(a), which shows that engaging in energy saving behaviours (denoted by behaviour) strengthens problem awareness, which further leads to more energy saving behaviours over time.



(a) A directed graph displaying feedback between problem awareness and engaging in energy saving behaviours. Note that this is not a DAG.



(b) A dynamic DAG, representing reciprocal causation between motivation and energy saving behaviours, obtained by rolling out the graph displayed in Figure 4(a). Note that the variables are now indexed by time.

Figure 4. DAGs encode feedback by taking time explicitly into account thereby allowing for underlying dynamics to be studied.

Causal Discovery

The examples and systematic approach we present in this paper assumes that the theories underlying the effect of an intervention on household energy usage is sufficiently developed to guide experts to draw a causal graph. However, in cases when when there is no clear theory, causal discovery algorithms can be used to explore the underlying causal graph structure (i.e., a DAG) in a data driven manner. This may inspire novel theorizing on how intervention programmes affect energy saving behaviours, that can next be tested on a new dataset. Causal discovery based on graphical causal models use the notion of conditional independence, and *d-separation* in particular, to learn the underlying DAG structure. It is beyond the scope of this paper to discuss causal discovery in detail, and interested readers are guided to Eberhardt (2016) for a brief introduction, Spirtes and Zhang (2016) for a review, and Spirtes et al. (2000) for a detailed presentation of causal discovery algorithms.

Limitations of DAGs

Graphical causal models, and DAGs in particular, are a tool to explicate causal assumptions and systematically understand sources of bias when RCTs are not feasible. However, there are limitations to using DAGs to evaluate effects of intervention programmes on household energy saving behaviour (Elwert, 2013). Firstly, drawing a DAG that adequately captures the theory describing how the intervention programme affects behaviour implies that researcher should have a clear theory on which factors may affect intervention effects. In addition, in the household energy domain, experts from multiple disciplines often work together, and incorporating their theories in one DAG can be challenging (Shrier & Platt, 2008). Furthermore, causal inference based on DAGs assumes that all relevant common causes are known and measures. As such, the possibility of latent (hidden) confounders poses an additional problem to the causal effects estimated based on a DAG (Pearl, 2009). Finally, causal discovery methods cannot recover some important aspects of the underlying causal processes, such as the functional form of the relations (e.g. linear or non-linear) and interactions.

Discussion

The aim of this paper was to introduce the reader to graphical causal models, and DAGs in particular, to evaluate the effects of behavioural interventions on household energy savings when RCTs are not feasible. In the absence of randomisation, non-experimental designs such as quasi-experiments and living labs are commonly used. However, irrespective of the research design, careful examination of causal effects in the absence of randomisation requires a systematic approach to dealing with bias, and we propose DAGs as one such approach. In brief, DAGs can increase our confidence in the causal claims when non-experimental designs are used (Steiner, Kim, Hall, & Su, 2017).

A systematic approach to causal inference based on DAGs has several advantages. Firstly, graphs are an intuitive way of representing the causal processes underlying the effects of behavioural intervention programmes on energy saving behaviours. Secondly, by approaching bias systematically, interventions can be evaluated more carefully leading to greater confidence in causal claims. Finally, as these models emphasize the need to develop sound theory on how interventions affect energy saving behaviours, they improve our understanding of the process underlying the effects of intervention programmes on household energy saving behaviours. In addition, in cases when there is no clear theory, data-driven causal discovery algorithms can guide researchers towards generating plausible theories that can then be tested in follow-up research.

Graphical causal models such as DAGs benefit science as they lead to a better understanding of processes underlying the effects of intervention programmes, and identify potential biases that may affect the evaluation of the effects of such interventions. Moreover, they result in better input for policy makers as they ensure a more rigorous evaluation of intervention programmes. We hope that approaching causal inference formally using methods such as graphical causal models will lead to an improved design, rigorous evaluation, and a better understanding of the processes underlying intervention programmes.

References

Abrahamse, W., Steg, L., Vlek, C. A., & Rothengatter, T. (2005). A review of intervention studies aimed at household energy conservation. *Journal of environmental psychology*, *25*(3), 273–291. doi: 10.1016/j.jenvp.2005.08.002

Albers, C. J. (2015). Dutch research funding, gender bias, and Simpson's paradox. *Proceedings of the National Academy of Sciences*, *112*(50). doi: 10.1073/pnas.1518936112

Allcott, H., & Mullainathan, S. (2010). Behavior and energy policy. *Science*, *327*(5970), 1204–1205. doi: 10.1126/science.1180775

Carey, T. A., & Stiles, W. B. (2016). Some problems with randomized controlled trials and some viable alternatives. *Clinical psychology & psychotherapy*, *23*(1), 87–95. doi: 10.1002/cpp.1942

Cole, S. R., Platt, R. W., Schisterman, E. F., Chu, H., Westreich, D., Richardson, D., & Poole, C. (2010). Illustrating bias due to conditioning on a collider. *International journal of epidemiology*, *39*(2), 417–420. doi: 10.1093/ije/dyp334

Cook, T. D., Campbell, D. T., & Shadish, W. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin Boston.

Deaton, A., & Cartwright, N. (2016). *Understanding and misunderstanding randomized controlled trials* (Working Paper No. 22595). National Bureau of Economic Research. doi: 10.3386/w22595

Eberhardt, F. (2016). Introduction to the foundations of causal discovery. *International Journal of Data Science and Analytics*, 1–11. doi: 10.1007/s41060-016-0038-6

Elwert, F. (2013). *Handbook of causal analysis for social research* (S. L. Morgan, Ed.). Dordrecht: Springer Netherlands.

Elwert, F., & Winship, C. (2014). Endogenous selection bias: The problem of conditioning on a collider variable. *Annual Review of Sociology*, *40*, 31–53. doi: 10.1146/annurev-soc-071913-043455

Frederiks, E. R., Stenner, K., Hobman, E. V., & Fischle, M. (2016). Evaluating energy behavior change programs using randomized controlled trials: Best practice guidelines for policymakers. *Energy Research & Social Science*, *22*, 147–164. doi: 10.1016/j.erss.2016.08.020

Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal diagrams for epidemiologic research. *Epidemiology*, 37–48. doi: 10.1097/00001648-199901000-00008

IPCC. (2014). *Climate change 2014–impacts, adaptation and vulnerability: Regional aspects*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Kalisch, M., Mächler, M., Colombo, D., Maathuis, M. H., & Bühlmann, P. (2012). Causal inference using graphical models with the R package pcalg. *Journal of Statistical Software*, *47*(11), 1–26. doi: 10.18637/jss.v047.i11

Karlin, B., Zinger, J. F., & Ford, R. (2015). The effects of feedback on energy conservation: A meta-analysis. *Psychological bulletin*. doi: 10.1037/a0039650

Kievit, R. A., Frankenhuis, W. E., Waldorp, L. J., & Borsboom, D. (2013). Simpson's paradox in psychological science: a practical guide. *Frontiers in Psychology*, *4*, 513. doi: 10.3389/fpsyg.2013.00513

Lilienfeld, S. O., McKay, D., & Hollon, S. D. (2018). Why randomised controlled trials of psychological treatments are still essential. *The Lancet Psychiatry*.

Pearl, J. (2009). *Causality: Models, reasoning and inference* (2nd ed.). New York, NY, USA: Cambridge University Press.

Pearl, J. (2014). Comment: Understanding simpson's paradox. *The American Statistician*, *68*(1), 8-13. doi: 10.1080/00031305.2014.876829

R Development Core Team. (2008). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org` (ISBN 3-900051-07-0)

Shadish, W. R., Clark, M. H., & Steiner, P. M. (2008). Can nonrandomized experiments yield accurate answers? a randomized experiment comparing random and nonrandom assignments. *Journal of the American Statistical Association*, *103*(484), 1334-1344. doi: 10.1198/016214508000000733

Shrier, I., & Platt, R. W. (2008). Reducing bias through directed acyclic graphs. *BMC medical research methodology*, *8*(1), 70. doi: 10.1186/1471-2288-8-70

Spirtes, P., Glymour, C. N., & Scheines, R. (2000). *Causation, prediction, and search*. Cambridge, MA, USA: MIT press.

Spirtes, P., & Zhang, K. (2016). Causal discovery and inference: concepts and recent methodological advances. *Applied Informatics*, *3*(1), 3. doi: 10.1186/s40535-016-0018-x

Steiner, P. M., Kim, Y., Hall, C. E., & Su, D. (2017). Graphical models for quasi-experimental designs. *Sociological Methods & Research*, *46*(2), 155-188. doi: 10.1177/0049124115582272

Textor, J., van der Zander, B., Gilthorpe, M. S., Liśkiewicz, M., & Ellison, G. T. (2016). Robust causal inference using directed acyclic graphs: the r package 'dagitty'. *International journal of epidemiology*, *45*(6), 1887–1894. doi: 10.1093/ije/dyw341

Vandenbroucke, J. P. (2008, 03). Observational research, randomised trials, and two views of medical science. *PLOS Medicine*, *5*(3), 1-5. doi: 10.1371/journal.pmed.0050067

Vine, E., Sullivan, M., Lutzenhiser, L., Blumstein, C., & Miller, B. (2014). Experimentation and the evaluation of energy efficiency programs. *Energy Efficiency*, *7*(4), 627–640. doi: 10.1007/s12053-013-9244-4

West, S. G. (2009). Alternatives to randomized experiments. *Current Directions in Psychological Science*, *18*(5), 299-304. doi: 10.1111/j.1467-8721.2009.01656.x