

Quasi-experimental causality in neuroscience and behavioural research

Ioana E. Marinescu^{1*}, Patrick N. Lawlor² and Konrad P. Kording^{3,4}

In many scientific domains, causality is the key question. For example, in neuroscience, we might ask whether a medication affects perception, cognition or action. Randomized controlled trials are the gold standard to establish causality, but they are not always practical. The field of empirical economics has developed rigorous methods to establish causality even when randomized controlled trials are not available. Here we review these quasi-experimental methods and highlight how neuroscience and behavioural researchers can use them to do research that can credibly demonstrate causal effects.

Behavioural research asks a broad range of questions, and most of them are of a causal nature¹. When we ask how a drug affects a patient, we want to know its causal effect: does it make the patient better? We do not want to ask the correlational question: does taking the drug correlate with well-being? Characteristics of the patient such as socioeconomic status may affect both the probability of being prescribed a drug and the patient's well-being. Similarly, in neuroscience, we have many causal questions. For example, we are interested in how one brain area affects another brain area, as opposed to how the two brain areas are correlated. In psychology, we ask which interventions improve people's thriving, again not to be confused with correlation (people with big yachts are happier, but see ref. ²). The primary goal of the bulk of scientific research is to ask how elements of a system causally affect other elements. Causality is at the heart of many questions in behaviour and neuroscience.

Ignoring the difference between correlation and causality frequently leads scientists to incorrect conclusions. In one notorious example from medicine, a correlational study suggested that hormone replacement therapy (HRT) may decrease the risk of cardiovascular disease in post-menopausal women³. A later randomized controlled trial, however, showed the opposite⁴ — that HRT actually led to worse cardiovascular outcomes. The discrepancy likely resulted from influences of socioeconomic status⁵; women with higher socioeconomic status were both more likely to receive HRT and to have better outcomes. In neuroscience, we have the same problem; many studies are inherently correlational due to the difficulty of controlling neurons and neural states. Reliably identifying causality without randomized experiments is difficult.

The crucial problem in causal inference is confounding. We would like to estimate the influence between two variables X and Y , but there may be other variables that affect both X and Y . If we set the value of X , as we do in a randomized experiment, there is no issue because the other variables can only affect Y . Running such experiments is the basis of the experimental method⁶, and allows for a direct reading out of causal effects. This has been done, often at great cost, for education (for example, The Perry Preschool Project and The Carolina Abecedarian Project), neuroscience (for example, optogenetics, slice stimulation) and clinical psychology (for example,

therapy comparison), and is frequently done in online user interactions (for example, A/B tests). In medicine, randomization is the gold standard and is called a randomized controlled trial. When we can set the relevant variables and randomize, answering causal questions meaningfully is far more straightforward.

But if we can only observe a system, then confounding is a serious problem; we can never know if an apparent interaction between X and Y is real or is confounded by the other variables. This is often the case for a number of reasons. First, there are many variables that we cannot easily set, for example, the activity of neurons somewhere in the brain. Second, setting variables is often expensive, for example, in the case of large clinical trials^{7–9}. Third, randomized experiments can be unethical, since they can force us to withhold the intervention that we believe to be best. When we cannot set all of the variables of interest, confounding is a serious issue that makes it hard to learn about the effect of one variable on another. Yet, most of the world's ever-growing data do not come from randomized experiments and we should not waste this data.

In response to the confounding problem in observational data, there are two important schools of thought. One school attempts to build large, complex models that observe and model all confounders. Such models thus assume that confounding is unlikely or impossible. There are many widely used methods that make this 'unconfoundedness' assumption in neural and behavioural research, for example, Granger causality¹⁰ (see Box 1 for others). However, unconfoundedness is rarely plausible as virtually all systems that we study have more variables of importance than we can realistically measure or model. A second school of thought that has arisen in response to the confounding problem focuses on quasi-experiments¹¹. Although we may not assume unconfoundedness in general, we may still be able to find variables in our data that are assigned in a way that is as good as random. This Perspective focuses on discussing this second way of thinking (see Box 2 for nomenclature) about causal inference.

This second school of thought mainly comes from econometrics, and over the past few decades has developed a number of ways in which meaningful causal estimates can be obtained without randomization. Economists were obtaining unreliable results based on correlational methods, so they decided to "take the con out of

¹Department of Social Policy and Practice, University of Pennsylvania, Philadelphia, PA, USA. ²Division of Neurology, Children's Hospital of Philadelphia, Philadelphia, PA, USA. ³Departments of Neuroscience and Bioengineering, Leonard Davis Institute, Warren Center for Network Science, Wharton Neuroscience Initiative, University of Pennsylvania, Philadelphia, PA, USA. ⁴Canadian Institute For Advanced Research, Toronto, Ontario, Canada.

*e-mail: ioma@upenn.edu

Box 1 | Overview of causal inference techniques already used in neuroscience

There are subfields of neuroscience that aspire to causal inference from observational data. Network neuroscience, for example, seeks to identify connectivity (often termed functional or effective connectivity) between brain regions using recorded time series from a variety of modalities (for example, functional magnetic resonance imaging, electroencephalography, spike trains)^{37,44–46}. This connectivity is sometimes interpreted as causal, but the validity of this interpretation depends on context, and a priori plausibility³⁵. Importantly, the techniques listed below generally assume a lack of confounding by unobserved variables. In our view, this seems unlikely given the small number of observed signals and the high dimensionality of the brain. In this box, we review the most prominent techniques used in these fields. Note that these models can overlap, and that we have only presented the essence of each.

Granger causal models. The core intuition of Granger causality is that causes temporally precede effects¹⁰. A variable X (a time series) is said to Granger-cause Y (also a time series) if earlier values of X and Y predict Y better than earlier values of Y alone. That is, if the history of X improves predictions of Y , this is evidence that X causes Y . Granger causality has been used extensively in network neuroscience^{47,48} and macroeconomics⁴⁹, but not without criticism³⁶.

State-space models. This is a broad family of models^{50,51} in which one represents a system with one or more ‘state’ variables to characterize ‘the way the system is’. State variables may or

may not be observed, typically evolve over time, and can be related to system inputs and output. For example, hippocampal neural activity could be a state variable that is affected/caused by experimental conditions and gives rise to (causes) a functional magnetic resonance imaging blood-oxygen-level dependent signal that is measured. States can be modelled as causally affecting one another as well^{52,53}. This family of models includes dynamic causal modelling^{52,54}, some types of point-process models⁵⁵ and others.

Structural equation models. This is a type of regression model with multiple equations²⁹. There can be multiple dependent variables and multiple independent variables. Dependent variables can also depend on other dependent variables. The dependencies between variables can, in some contexts, be given a causal interpretation. The parameters of these models are often found by regression approaches. Such models are also used in economics⁵⁶.

Bayesian networks. This is a type of model that includes variables and their statistical dependencies^{1,28}. In this framework, causality can be viewed as the probabilistic influence variable X has on variable Y after taking into account other variables in the network. It is said to be Bayesian because variables have prior and likelihood distributions, and other tools of Bayesian statistics can be used. This is also a broad family of models, and has been used in both network neuroscience^{38,57} as well as human causal learning^{58–61}.

econometrics¹² by developing better tools for causal inference. Some of these methods include the regression discontinuity design (RDD)^{13,14}, the difference-in-differences (DiD) approach¹¹ and instrumental variables¹¹. These techniques are standard in economics yet are rarely used in many branches of behavioural and neuroscience research (although see Box 1 and Discussion for causal inference techniques already used in neuroscience).

Here we review these alternatives to randomization. We take published examples, and explain the methods. For each method, we then sketch how it could be used more widely across behavioural and neuroscience research using existing and emerging data. By systematically replacing correlational techniques with causal techniques, economics went through what they call a credibility revolution. Perhaps as a result, empirical work in economics has progressively overtaken theoretical work both in terms of citations within economics¹⁵, and in terms of citations to economics papers made by articles in other fields¹⁶.

Regression discontinuity design

It is possible to approximate causality from a common property of decisions: thresholds for treatment. Treatment effects can often be estimated near thresholds, even without randomized experiments, because subjects near the threshold are similar. This approach was originally developed by Thistlethwaite, who was interested in the effect of academic recognition of student outcomes¹³. In that study, students with test scores above a certain threshold were given certificates of merit and public recognition. The students who received the certificate were clearly different from those that did not, for example, in intelligence and socioeconomic status, so it was not possible to simply ask how certificates of merit affected outcomes. However, as we approach the threshold score from either side, the students will become arbitrarily similar. This is because there is randomness in the exact score a student received due to, for example, question selection or their last night’s sleep. Thistlethwaite found that certificates of merit led to more future scholarships, but not to

differences in long-term career plans¹³. This strategy, which is based on the idea that samples just above and just below the threshold are nearly indistinguishable, is the basis of the RDD.

To perform the analysis, a regression of outcome as a function of the running variable (for example, test scores in the study by Thistlethwaite) is fit on both sides of the threshold. A causal effect would be manifested by a discontinuity between the regression line on the left and on the right of the threshold (see Fig. 1). This discontinuity in the outcome can only be from the treatment because no confounder is likely to have a discontinuity at exactly the same threshold (although it is standard to check this assumption). The RDD gives a meaningful and often unbiased estimate of the causal effect of the treatment in the vicinity of the threshold¹⁴.

It is important to be aware of the conditions needed for causal validity in RDD. Subjects must not be able to precisely control their score — and thus their treatment — for example, by working long enough hours to achieve exactly the score that will put them over the threshold. A test for this assumption was developed¹⁷, which looks for a discontinuity in the number (that is, density) of subjects on either side of the threshold. Moreover, subjects must not be able to override the threshold mechanism for selection. Sometimes so-called fuzzy RDD approaches can deal with the problem of treatments not being perfectly administered¹⁸. Best practices for implementing RDD can be found in refs ^{14,19}. Importantly, these methods allow checking whether the assumptions underlying RDD are valid. For example, potential confounders should not have a discontinuity at the threshold. While there are many statistical issues to consider for the RDD, there is an active community of practitioners furthering our already-strong understanding.

RDD should allow us to discover causal effects in many domains (see Table 1). Thresholds exist widely in human behaviour and neuroscience, and there are very few variables that subjects can noiselessly control. For example, in the field of neural theory, we might ask how neurons can estimate their causal effect on animal performance, which would allow them to ask whether a larger weight

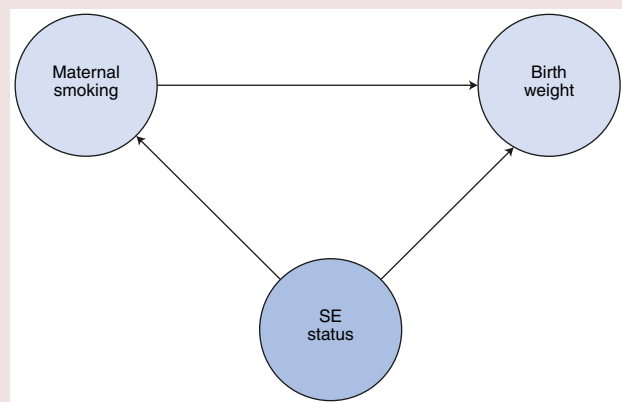
Box 2 | Introduction to graphical models in causal reasoning

When trying to infer causal effects, it is helpful to visually represent the variables under consideration and the relationships between them. Graphical models, widely used in computational fields, provide a way to do this by representing each relevant variable as a circle (with a label), and each putative statistical relationship between two variables as a line connecting them. Causal relationships are represented by arrows, which can be unidirectional or bidirectional. The variables included in a graphical model should be all relevant independent and dependent variables, as well as confounding variables that may affect the independent and dependent variables. Framed this way, the goal of causal inference becomes clearer: to estimate the strength and direction of the statistical relationships (lines/arrows) between variables after maximally accounting for the important components in the system. Actually estimating the statistical relationships depends on the specifics of the proposed model.

Importantly, we should not assume that the variables we use in a graphical model can be experimentally controlled. We should, therefore, distinguish between observed variables and controlled (or set) variables. Observed variables always have the possibility of being affected by unknown and unmodelled confounding variables, whereas controlled variables are immune to this problem. Following Pearl¹, we use the notation of $\text{do}(X)$ to indicate an experimentally controlled variable, and the unqualified X to indicate a variable that is simply observed.

Consider an example addressed in the main text, in which we seek to find out whether maternal smoking affects a child's birth weight. To form a graphical model of this scenario, we

would specifically model maternal smoking and birth weight as variables in the system. Because we believe that smoking may influence birth weight, we would draw an arrow that points from maternal smoking to birth weight. We would also want to account for confounding factors, such as socioeconomic (SE) status; these factors may influence both a mother's smoking as well as birth weight. Socioeconomic status should also be specifically modelled in the graphical model, and we should draw arrows from socioeconomic status to both maternal smoking and birth weight. Furthermore, because ethically we cannot randomize maternal smoking, we cannot use the $\text{do}()$ notation. The graphical model of this simplified system is shown below.



would be better. The translation of neural drive to spiking has a firing threshold, which could allow neurons to estimate their causal effect²⁰. There are countless possibilities to expand on the small set of current applications, for example, refs^{13,21}. RDD thus promises to be useful for nearly every subfield of behavioural research and neuroscience, ranging from education and medicine all the way to neural theory.

Difference-in-differences

Another approach for approximating causality is to look for temporal trends in treated versus untreated subjects, even if they were not randomly assigned. The core idea of this approach is to use longitudinal data for two groups where only one is treated, but where the two groups are similarly affected by extraneous factors. For example, ref.²² investigated the effect of academic year length on student outcomes. It exploited a transient reduction in school year length that occurred in some but not all German states. Thus, it was possible to compare outcomes between short-school-year states (the treatment group) and a regular-school-year state (the control group) to measure the effect of the shortened school year. Grade repetition increased in the short-school-year states relative to the regular-school-year state after the short school year was introduced. The length of the school year was thus found to have a causal effect on repeating grades.

To perform the analysis, the temporal evolution of the outcome is measured for both the treated and the untreated group. This, in a way, generalizes the idea of baseline-controlled or two-factorial designs sometimes used in clinical trials. A quantification of the temporal difference between the treated and the untreated group then allows the treatment effect to be estimated (see Fig. 2).

The DiD approach naturally comes with its own assumptions and caveats, many of which we can explicitly test. Most importantly, it assumes that the two groups are chosen such that they are similarly affected by relevant and perhaps unmeasured factors: this

is the common trends assumption. In the above example, the two groups of German states should be similarly affected by the economic context, other policy changes and so on. One way to provide support for this assumption is to check that trends in the outcome before the new treatment are parallel. The groups should also be stable in composition (for example, percentage of women in each group) over the period of comparison. Extensions such as nonlinear DiD have also been developed²³. Best practices for DiD can be found in ref.¹¹.

DiD approaches should also be broadly applicable in behavioural science and neuroscience (see Table 2). Many, if not most, variables in neuroscience and behaviour are measured over time. And many interventions affect some people or neurons (the treatment group) but not others (the control group). The DiD approach thus promises to be useful across most subdisciplines that deal with behaviour.

Instrumental variables

A third common approach for quasi-experimental causal inference is instrumental variables²⁴. With this approach, we seek to identify variables Z ('instruments') that causally affect the independent variable of interest X , but only causally affect the dependent variable Y through X (Fig. 3). For example, ref.²⁵ sought to ask how maternal smoking affects birth weight. We should expect heavy confounding as, for example, low socioeconomic status may affect both smoking and health. Instead, the authors leveraged tobacco taxes as an instrument, which arguably affects smoking but does not directly affect birth weight. Differences in tobacco taxes across years and across states could then be exploited to estimate the causal effect of smoking on birth weight. They found that maternal smoking decreased birth weight by between 300 and 600 g.

To perform an instrumental variable analysis, we first identify the independent and dependent variables. Next we find, through an understanding of the system, another variable that can serve as an

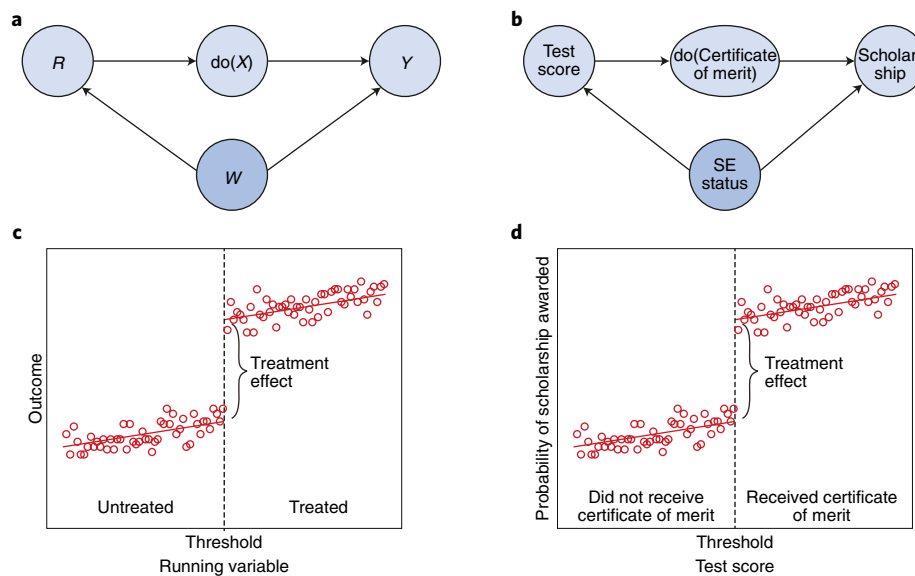


Fig. 1 | Regression discontinuity design. **a**, Graphical model of the RDD approach. W represents confounding variables; R is the running variable, which determines the treatment along with the threshold; X is the treatment (independent variable), which is either administered ($do(X)$) or not administered ($do(\text{not } X)$) depending on R ; and Y is the outcome (dependent variable) of interest. **b**, Graphical model representing this analysis performed in ref. ¹³. Socioeconomic (SE) status (for example) is likely to affect both test score and the probability of receiving a scholarship. Test score determines whether a certificate of merit is awarded, which in turn affects the probability of receiving a scholarship. **c**, Schematic of a RDD analysis. The treatment is only administered if the running variable is above the threshold. The outcome (y axis) is plotted as a function of a running variable (x axis). The magnitude of the treatment effect, the difference in outcome at the threshold, is estimated using regression. **d**, Schematic representing the analysis performed in ref. ¹³. Academic outcome (probability of scholarship) is plotted as a function of test score, and a discontinuity is seen at the cut-off for receiving a certificate of merit. Note that this figure is stylized and does not use the data used in the original analysis; it is intended only to demonstrate the approach.

instrument (taxes in the above example) that only affects the independent variable. Next, we build a predictive model of the treatment X based on the instrument Z (first-stage regression). And then we use this prediction in a second-stage regression to quantify the causal effect of treatment (for example, smoking) on the dependent variable (birth weight in the above example). The essence of this approach is that it identifies changes in the dependent variable that occur as a result of varying the independent variable; the instrument can be viewed as a rudimentary experimental manipulation of the independent variable. Simply regressing the dependent variable on the independent variable would be confounded by, for example, socioeconomic status in the above example. Such instrumental variable approaches allow yet one more way to meaningfully deal with unobserved confounders.

The instrumental variable approach also has its assumptions and caveats. The most important assumption is the exclusion restriction: the instrument (for example, tobacco taxes) should affect the dependent variable (for example, birth weight) only through its effect on smoking. That is, we should be able to exclude that the instrument affects the outcome other than through the independent variable. The exclusion restriction is not directly testable and must therefore be assessed on plausibility grounds given what we know about the phenomenon at hand. Furthermore, the instrument should not be too weakly correlated with the independent variable of interest to produce useful estimates. The F test for the first stage can measure how strong the instrument is²⁶. Best practices and further discussion of instrumental variable can be found in ref. ¹¹

Instrumental variable approaches should also be broadly applicable in behavioural science (see Table 3). Typical experiments have many variables that are not experimentally randomized, yet affect the system. For example, metabolism affects neural activity, which

affects behaviour; markers of metabolism could therefore be viewed as instruments to ask how neural activity gives rise to behaviour. Many such variables are random with respect to behaviour, and could just as easily be viewed as instruments. Even standard techniques such as optogenetics may be better viewed as instruments²⁷. Optogenetics does not precisely set neural activity; it only affects it. Hence, it may be useful to view optogenetic stimulation of brain region X as an instrumental variable, and then use that model to ask how brain region X affects region Y . More generally, it may be possible to create biological constructs with instrumental variable-like properties; for example, a molecular construct that inactivates individual neurons at random times. In summary, instrumental variables are an approach to get good results by using existing, non-experimental randomization.

Discussion

Here we have argued that understanding causal effects is the goal of the bulk of both behavioural science and neuroscience, and that these fields need to adopt better techniques for making causal inferences. We have reviewed three prominent quasi-experimental approaches developed in economics, explained their application, and suggested ways that they might be applied in a number of examples using existing data. These techniques promise to move our data analysis towards a causal understanding. We chose three particular techniques — RDD, DiD and instrumental variables — but many other techniques for estimating causal influence exist. For example, Bayesian networks²⁸ and structural equations²⁹ can be used to model networks of relevant variables and to estimate causal relationships between them. Propensity score matching estimates causal effects between treated and untreated subjects by adjusting for observed confounders that predict treatment^{30–32}. Other techniques use noise distributions to estimate the direction of causal

Table 1 | Possible applications of RDD in neuroscience and behavioural research

Area	Question	Running variable	Threshold	Outcome variable
Education	How much does enrichment help?	Test scores used for enrichment programme	Minimum test score	Education outcome, income
Medicine	How much does blood pressure medication help?	Blood pressure	Treatment guideline	Death by cardiovascular disease
Counselling	How many people should receive depression treatment?	Risk score	Enrolment threshold	Mental health
Advertising	How much does an advertisement affect consumer behaviour?	Affinity score	Money limit	Sale of product
Neural data science	What are the neural requirements for movement?	Neural drive	Firing threshold	Activity of a downstream neuron or muscle
Neural theory	How much would a larger synaptic weight increase reward-seeking behaviour?	Neural drive	Firing threshold	Behavioural change

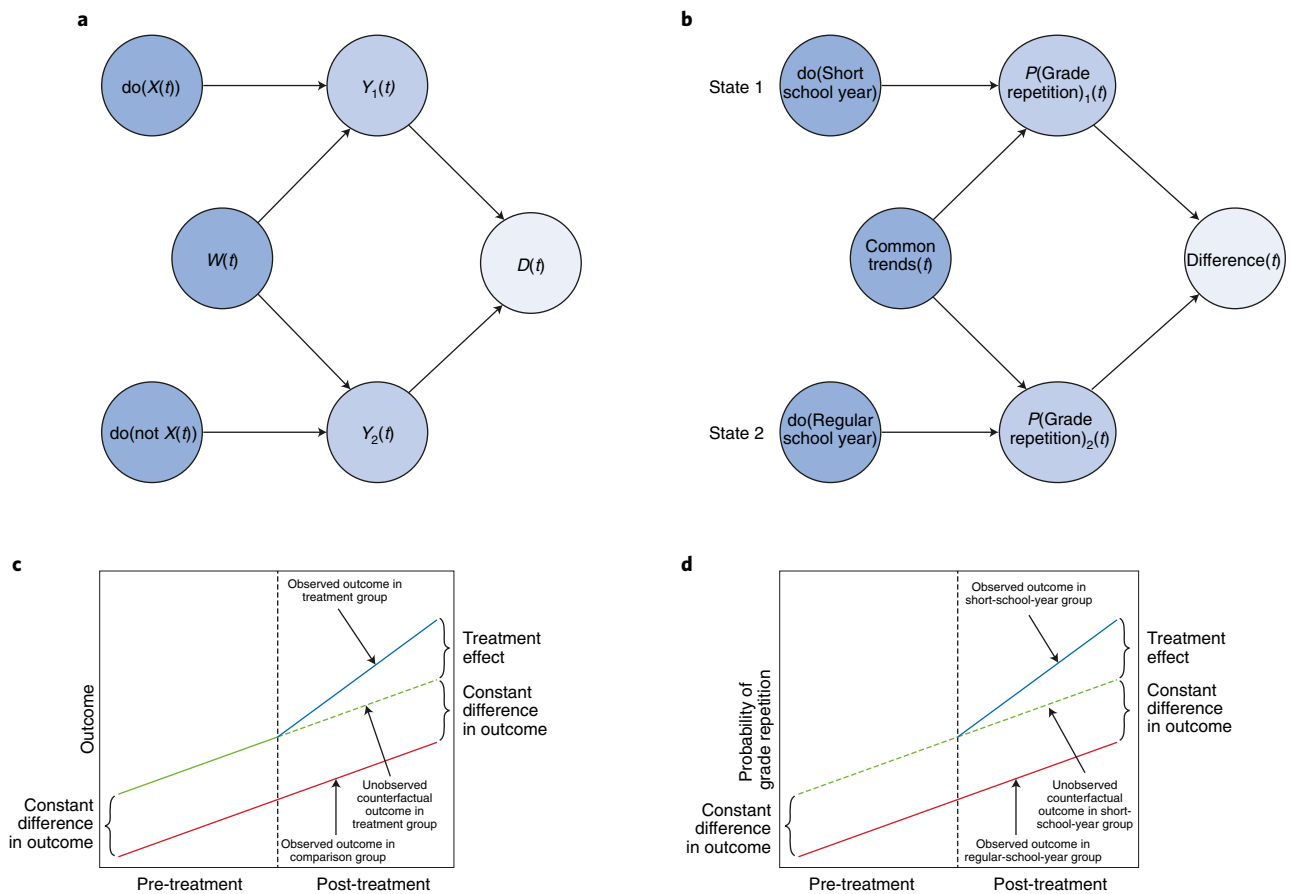


Fig. 2 | Difference-in-differences. **a**, Graphical model for the DiD approach. All variables are considered as a function of time, t . W represents confounding variables; X is the treatment (independent variable), which is administered ($do(X)$) to population 1, and not administered ($do(not X)$) to population 2; Y_1 and Y_2 are the outcomes (dependent variables) for populations 1 and 2, respectively; D is the difference between Y_1 and Y_2 and is tracked over time. **b**, Graphical model representing this analysis performed in ref. ²². Common trends such as federal taxes and economic conditions are likely to affect the two states similarly. The short school year is implemented only in one state. The difference in outcome is calculated from the two states' outcomes. P , probability. **c**, Schematic of a DiD analysis. The trend of two groups, treated and untreated, is plotted as a function of time. Before the treatment, the trends of the two groups should be parallel (a constant DiD). The treatment effect is estimated by the degree to which the trends diverge after the treatment is administered. **d**, Schematic figure representing the analysis performed in ref. ²². Outcome (probability of grade repetition) is plotted as a function of time, before and after the implementation of the short school year in some states. The difference between state outcomes changes after the change in school year (that is, there is an increase in difference in differences). Note that this figure is stylized and does not use the data used in the original analysis; it is intended only to demonstrate the approach.

Table 2 | Possible applications of the DiD approach

Field	Question	Comparison	Outcome measured over time
Educational policy	How do smartphones affect middle-school students?	Two nearby school districts before and after a new smartphone policy change	Standardized test scores, disciplinary action
Rehabilitation	How well does rehab work?	Affected limb and unaffected limb before and after rehab	Strength, coordination scores
Neurology	What are the effects of new brain lesions in multiple sclerosis?	Before and after unilateral lesion	Strength, coordination scores
Public relations	Do ads have negative side effects?	In- versus outside of target area, before and after the start an ad campaign	Attitudes towards ads

influence³³. Synthetic controls may be a valuable way to construct better control/comparison groups for case control and DiD studies³⁴. In this Perspective, however, we have focused on quasi-experimental approaches as these approaches readily allow dealing with unobserved confounders that make causal inference difficult in behavioural science as well as neuroscience.

Neuroscience and psychology do have a history of using techniques that attempt to recover causal influences from data. Network neuroscience has used a large suite of approaches (see Box 1) with the goal of deciphering the complex networks of the brain. This is an important line of work, but whether these techniques actually recover true causal effects in this context remains an open question^{35,36}. Many challenges exist, such as possible omitted variables and the difficulty of modelling information transformation between brain regions^{35,37}. Although some of the techniques we present could be applied towards this purpose, we view the RDD, DiD and instrumental variables as more general approaches that exploit different aspects of data than network approaches. The RDD exploits thresholds, which is clearly different from the network approaches. The DiD approach exploits common trends even when confounders are not always identifiable, whereas network approaches generally are sensitive to omitted confounders³³. The instrumental variables approach identifies non-experimental randomization, and although it could be incorporated into network approaches, we believe that

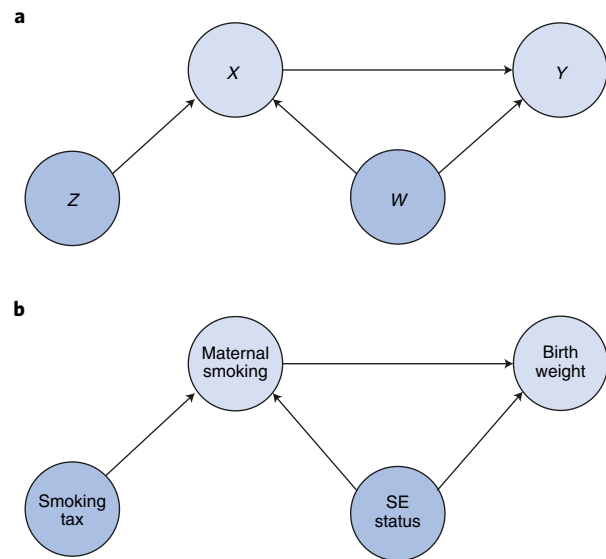


Fig. 3 | Instrumental variables. **a**, Graphical model for the instrumental variables approach. *W* represents confounding variables; *X* is the independent variable; *Y* is the outcome (dependent variable); *Z* is the instrument, which only affects *Y* through its effect on *X*. **b**, Graphical model representing the analysis performed in ref. ²⁵. Maternal smoking is thought to affect birth weight. But socioeconomic (SE) status (for example) likely affects both a mother’s decision to smoke as well as the child’s birth weight. A tax on cigarette smoking could affect maternal smoking but is unlikely to directly influence the birth weight, except through an effect on maternal smoking. Such a tax is therefore a good instrument to examine the effect of smoking on birth weight without being confounded by socioeconomic status.

its use is far more general. We therefore believe that the techniques discussed in this Perspective widen the scope of data available for causal analysis in neuroscience and behavioural science.

Techniques for quasi-experimental causal inference are ripe for application in behavioural science and neuroscience. They could fruitfully be applied to existing laboratory data, such as neuroimaging, virtual reality behaviour or neural spike recordings. This may allow us to extract more valuable information from these data. But these techniques also make it possible to perform credible analyses of the kind of observational data offered by the information age³⁹ that is much cheaper and much more common than laboratory data. Thresholds exist everywhere, in online systems⁴⁰, economic activity (for example, tax notches, see ref. ⁴¹), but especially in medicine²¹, making RDD, with its clean treatment of confounders, an invaluable tool. Wherever parallel trends exist, DiD promises to give our analyses better controls. And identifying valid instru-

Table 3 | Possible applications of instrumental variable approach

Area	Question	Independent variable <i>X</i>	Dependent variable <i>Y</i>	Instrument <i>Z</i>
Medicine	Does a medication affect patient outcomes?	Medication use	Patient outcome	Hospital rules about pharmaceutical reps
Neuroscience	How does brain region A affect region B?	Brain region A activity	Brain region B activity	Diffuse optogenetic stimulation of brain region A
Behavioural health	Does alcohol consumption make you a bad parent?	Alcohol consumption	Parenting license exam	Alcohol taxes
Education	Does having more disposable income improve educational outcomes?	Income	Educational outcome	Income tax cut

ments on independent variables of interest will help us to tease apart causal relationships with instrumental variables. Every neuroscientist and behavioural scientist should become familiar with these techniques.

In the 1980s, it became obvious in economics that the typical correlational findings were not overly indicative of real causal effects. This led the field of econometrics to decide to work towards methods that allow the quantification of causality⁴². In the following decades, causal inference improved massively and today the bulk of top economics papers uses standard causal inference strategies⁴².

Neuroscience and behavioural science have the same problem: we write stories about causality in behaviour and brains based on correlational data, whereas we need techniques that can reliably demonstrate causal effects. Many techniques currently used in neuroscience may actually be misleading us³⁶ because we misunderstand whether they measure causal effects. Furthermore, a focus on causal effects should help us to focus on which effects are worth caring about, such as behaviour⁴³. This understanding of the deeply problematic basis of this kind of inference is slowly taking hold in the community. To lead to a deeper understanding of minds and brain, we need to take the causal questions seriously. We can only do so by applying techniques that allow us to answer those questions.

Received: 18 March 2018; Accepted: 2 October 2018;
Published online: 26 November 2018

References

- Pearl, J. *Causality* (Cambridge Univ. Press, New York, 2009).
- The Notorious B.I.G. *Mo' Money Mo' Problems* (Bad Boy Records, 1997).
- Grodstein, F. et al. A prospective, observational study of postmenopausal hormone therapy and primary prevention of cardiovascular disease. *Ann. Intern. Med.* **133**, 933–941 (2000).
- Manson, J. E. et al. Estrogen plus progestin and the risk of coronary heart disease. *N. Engl. J. Med.* **349**, 523–534 (2003).
- Humphrey, L. L., Chan, B. K. & Sox, H. C. Postmenopausal hormone replacement therapy and the primary prevention of cardiovascular disease. *Ann. Intern. Med.* **137**, 273–284 (2002).
- Greenland, S. Randomization, statistics, and causal inference. *Epidemiology* **1**, 421–429 (1990).
- Ismail-Beigi, F. et al. Effect of intensive treatment of hyperglycaemia on microvascular outcomes in type 2 diabetes: an analysis of the ACCORD randomised trial. *Lancet* **376**, 419–430 (2010).
- Officers, T. A. Major outcomes in high-risk hypertensive patients randomized to or calcium channel blocker vs diuretic. *J. Am. Med. Assoc.* **288**, 2981–2997 (2002).
- Group, S. R. A randomized trial of intensive versus standard blood-pressure control. *N. Engl. J. Med.* **373**, 2103–2116 (2015).
- Granger, C. W. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**, 424–438 (1969).
- Angrist, J. D. & Pischke, J.-S. *Mostly Harmless Econometrics: An Empiricist's Companion* (Princeton Univ. Press, Princeton, 2008).
- Leamer, E. E. Let's take the con out of econometrics. *Am. Econ. Rev.* **73**, 31–43 (1983).
- Thistlethwaite, D. L. & Campbell, D. T. Regression-discontinuity analysis: an alternative to the ex-post facto experiment. *J. Educ. Psychol.* **51**, 309–317 (1960).
- Imbens, G. W. & Lemieux, T. Regression discontinuity designs: a guide to practice. *J. Econom.* **142**, 615–635 (2008).
- Angrist, J., Azoulay, P., Ellison, G., Hill, R. & Lu, S. F. Economic research evolves: fields and styles. *Am. Econ. Rev.* **107**, 293–297 (2017).
- Angrist, J., Azoulay, P., Ellison, G., Hill, R. & Lu, S. F. *Inside Job or Deep Impact? Using Extramural Citations to Assess Economic Scholarship* (National Bureau of Economic Research, 2017).
- McCrary, J. Manipulation of the running variable in the regression discontinuity design: a density test. *J. Econom.* **142**, 698–714 (2008).
- Trochim, W. M. *Research Design for Program Evaluation: The Regression-Discontinuity Approach* (Sage Publications, Beverly Hills, 1984).
- Jacob, R., Zhu, P., Somers, M. A., & Bloom, H. A practical guide to regression discontinuity. *MDRC* <https://www.mdrc.org/publication/practical-guide-regression-discontinuity> (2012).
- Lansdell, B. & Kording, K. Spiking allows neurons to estimate their causal effect. Preprint at bioRxiv <https://doi.org/10.1101/253351> (2018).
- Moscoe, E., Bor, J. & Bärnighausen, T. Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice. *J. Clin. Epidemiol.* **68**, 122–133 (2015).
- Pischke, J. S. The impact of length of the school year on student performance and earnings: evidence from the German short school years. *Econ. J.* **117**, 1216–1242 (2007).
- Athey, S. & Imbens, G. W. Identification and inference in nonlinear difference-in-differences models. *Econometrica* **74**, 431–497 (2006).
- Angrist, J. D. et al. Identification of causal effects using instrumental variables. *J. Am. Stat. Assoc.* **91**, 444–455 (1986).
- Evans, W. N. & Ringel, J. S. Can higher cigarette taxes improve birth outcomes? *J. Public. Econ.* **72**, 135–154 (1999).
- Stock, J. H. & Yogo, M. *Testing for Weak Instruments in Linear IV Regression*. (National Bureau of Economic Research, Cambridge, 2002).
- Li, X., Yamawaki, N., Barrett, J. M., Körding, K. P. & Shepherd, G. Scaling of optogenetically evoked signaling in a higher-order corticocortical pathway in the anesthetized mouse. *Front. Syst. Neurosci.* **12**, 16 (2018).
- Koller, D. & Friedman, N. *Probabilistic Graphical Models: Principles and Techniques* (MIT Press, Cambridge, 2009).
- Ullman, J. B., & Bentler, P. M. in *Handbook of Psychology* 2nd edn (eds Schinka, J. A. & Velicer, W. F.) Ch. 23 (John Wiley & Sons, Hoboken, 2012).
- Dehejia, R. H. & Wahba, S. Propensity score-matching methods for nonexperimental causal studies. *Rev. Econ. Stat.* **84**, 151–161 (2002).
- Rosenbaum, P. R. & Rubin, D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55 (1983).
- Imbens, G. W. & Rubin, D. B. *Causal Inference in Statistics, Social, and Biomedical Sciences* (Cambridge Univ. Press, New York, 2015).
- Hoyer, P. O., Janzing, D., Mooij, J. M., Peters, J. & Schölkopf, B. Nonlinear causal discovery with additive noise models. *Adv. Neural Inf. Process. Syst.* **21**, 689–696 (2009).
- Abadie, A., Diamond, A. & Hainmueller, J. Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. *J. Am. Stat. Assoc.* **105**, 493–505 (2010).
- Ramsey, J. D. et al. Six problems for causal inference from fMRI. *NeuroImage* **49**, 1545–1558 (2010).
- Jonas, E. & Kording, K. P. Could a neuroscientist understand a microprocessor? *PLoS Comput. Biol.* **13**, e1005268 (2017).
- Valdes-Sosa, P. A., Roebroeck, A., Daunizeau, J. & Friston, K. Effective connectivity: influence, causality and biophysical modeling. *NeuroImage* **58**, 339–361 (2011).
- Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M. & Beckmann, C. F. et al. Network modelling methods for FMRI. *NeuroImage* **54**, 875–891 (2011).
- Kwak, H., Lee, C., Park, H. & Moon, S. What is Twitter, a social network or a news media? In *Proc. 19th International Conference on World Wide Web* 591–600 (ACM, 2010).
- Bem, J. Using match confidence to adjust a performance threshold. Google Patent 7346615 (2008).
- Slemrod, J. Buenas notches: lines and notches in tax system design. *EJ Tax Res.* **11**, 259–283 (2013).
- Angrist, J. D. & Pischke, J.-S. The credibility revolution in empirical economics: how better research design is taking the con out of econometrics. *J. Econ. Perspect.* **24**, 3–30 (2010).
- Krakauer, J. W., Ghazanfar, A. A., Gomez-Marín, A., MacIver, M. A. & Poeppel, D. Neuroscience needs behavior: correcting a reductionist bias. *Neuron* (in the press).
- Pillow, J. W. et al. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* **454**, 995–999 (2008).
- Stevenson, I. H. & Körding, K. P. On the similarity of functional connectivity between neurons estimated across timescales. *PLoS ONE* **5**, e9206 (2010).
- Sakkalis, V. Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG. *Comput. Biol. Med.* **41**, 1110–1117 (2011).
- Bressler, S. L. & Seth, A. K. Wiener–Granger causality: a well established methodology. *NeuroImage* **58**, 323–329 (2011).
- Ding, M., Chen, Y. & Bressler, S. L. in *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications* (eds Schelter, B., Winterhalder, M. & Timmer, J.) 335–368 (Wiley-VCH, Weinheim, 2006).
- Hiemstra, C. & Jones, J. D. Testing for linear and nonlinear Granger causality in the stock price — volume relation. *J. Finance* **49**, 1639–1664 (1994).
- Chen, Z. *Advanced State Space Methods for Neural and Clinical Data* (Cambridge Univ. Press, Cambridge, 2015).
- Shumway, R. H. & Stoffer, D. S. in *Time Series Analysis and its Applications* (Shumway, R. H. & Stoffer, D. S.) 319–404 (Springer, New York, 2011).
- Friston, K. J., Harrison, L. & Penny, W. Dynamic causal modelling. *NeuroImage* **19**, 1273–1302 (2003).

53. Semedo, J., Zandvakili, A., Kohn, A., Machens, C. K. & Byron, M. Y. Extracting latent structure from multiple interacting neural populations. *Adv. Neural Inf. Process. Syst.* **27**, 2942–2950 (2014).
54. Daunizeau, J., David, O. & Stephan, K. E. Dynamic causal modelling: a critical review of the biophysical and statistical foundations. *NeuroImage* **58**, 312–322 (2009).
55. Latimer, K. W., Yates, J. L., Meister, M. L. R., Huk, A. C. & Pillow, J. W. Single-trial spike trains in parietal cortex reveal discrete steps during decision-making. *Science* **349**, 184–187 (2015).
56. Nevo, A. & Whinston, M. D. Taking the dogma out of econometrics: structural modeling and credible inference. *J. Econ. Perspect.* **24**, 69–82 (2010).
57. Song, L., Kolar, M. & Xing, E. P. Time-varying dynamic bayesian networks. *Adv. Neural Inf. Process. Syst.* **22**, 1732–1740 (2009).
58. Goodman, N. D., Ullman, T. D. & Tenenbaum, J. B. Learning a theory of causality. *Psychol. Rev.* **118**, 110–119 (2011).
59. Gopnik, A. et al. A Theory of causal learning in children: causal maps and Bayes nets. *Psychol. Rev.* **111**, 3–32 (2004).
60. Gopnik, A. & Tenenbaum, J. B. Bayesian networks, Bayesian learning and cognitive development. *Dev. Sci.* **10**, 281–287 (2007).
61. Körding, K. P. et al. Causal inference in multisensory perception. *PLoS ONE* **2**, e943 (2007).

Competing interests

The authors declare no competing interests.

Additional information

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence should be addressed to I.E.M.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© Springer Nature Limited 2018

Reproduced with permission of copyright owner. Further reproduction prohibited without permission.