Received: 12 January 2022 | Accepted: 16 May 2022

DOI: 10.1111/add.15972

METHODS AND TECHNIQUES

ADDICTION SSA

Causal inference with observational data in addiction research

Gary C. K. Chan[1] | Carmen Lim[1] | Tianze Sun[1] | Daniel Stjepanovic[1] | Jason Connor[1,2] | Wayne Hall[1] | Janni Leung[1]

[1]National Centre for Youth Substance Use Research, University of Queensland, Brisbane, QLD, Australia

[2]Discipline of Psychiatry, Faculty of Medicine, University of Queensland, Brisbane, QLD, Australia

Correspondence
Gary C. K. Chan, National Centre for Youth Substance Use Research, The University of Queensland, Brisbane, QLD 4072, Australia.
Email: c.chan4@uq.edu.au

Funding information
Department of Health, Australian Government - National Health and Medical Research Council APP1176137.

Abstract

Randomized controlled trials (RCTs) are the gold standard for making causal inferences, but RCTs are often not feasible in addiction research for ethical and logistic reasons. Observational data from real-world settings have been increasingly used to guide clinical decisions and public health policies. This paper introduces the potential outcomes framework for causal inference and summarizes well-established causal analysis methods for observational data, including matching, inverse probability treatment weighting, the instrumental variable method and interrupted time-series analysis with controls. It provides examples in addiction research and guidance and analysis codes for conducting these analyses with example data sets.

KEYWORDS
Causal inference, instrumental variable, interrupted time-series analysis, inverse probability treatment weighting, matching, propensity score

INTRODUCTION

Randomized controlled trials (RCTs) are the gold standard design for establishing a causal relationship between a treatment and an outcome. In RCTs, participants are randomized to one or more treatments and control (placebo) conditions. Randomization ensures that all measured and unmeasured variables are equally distributed across conditions, allowing the isolation of the treatment effect. RCTs are common in experimental and clinical research, but disadvantages include a lack of generalizability beyond the trial population, low statistical power for rare health conditions and high financial and time costs. Further, RCTs are not always feasible for some fields or research questions, such as addiction epidemiology and policy evaluation, due to ethical (e.g. testing the effect of smoking on cancer) and logistic consideration. Thus, addiction research often relies upon observational data. The importance of using observational data from real-world settings for causal inference has been increasingly recognized in health and medical research [1]. For example, in the United States, the 21st Century Cures Act supports the use of real-world evidence, including data from prospective and retrospective studies, when making health-care decisions and approving drugs [2].

In this paper, we provide an overview of established causal inference methods for non-randomized observational data [3] that is tailored for applied researchers with examples in substance use research. We will use the terms 'exposure', 'treatment' and 'intervention' interchangeably. We only focus upon research design with one treatment and one control condition. However, it should be noted that the ideas introduced in this paper can be extended to design with multiple treatments. This paper is not intended to be a comprehensive technical overview of the causal inference literature. Rather, it introduces researchers to several well-developed causal inference methods and orientates readers to the broader literature on causal inference. This paper is structured into five parts: (i) potential outcomes and counterfactual framework, (ii) matching, (iii) inverse probability treatment weighting, (iv) instrumental variable method and (v) interrupted time-series analysis.

PART I: POTENTIAL OUTCOMES FRAMEWORK

The definition of a causal effect has been formalized in the Rubin causal model with potential outcomes [4, 5]. For an individual *i*, the

causal effect of a treatment is defined as the difference between two potential outcomes: the outcome that would have been observed if the individual were exposed to treatment (denoted as Y_i^1, where i represented the i^{th} individual in the sample) and the outcome that would have been observed if the individual were not exposed to the treatment (denoted as Y_i^0). For example, in evaluating the effect of smoking on cancer, this would be equivalent to comparing the cancer outcome (yes/no) for an individual if this individual had smoked and if this individual had not smoked, with all other factors remaining the same. Only one of these outcomes can be observed; the other is a counterfactual. Therefore, the individual causal effect cannot be estimated. However, under several causal assumptions, as described below, it is possible to estimate the average treatment effect (ATE) by aggregating observations from a group of individuals who are representative of the target population. The ATE is defined as the average difference between Y_i^1 (what would have happened to the individual given exposure) and Y_i^0 (what would have happened given no exposure; i.e. $E\left(Y_i^1 - Y_i^0\right)$, where the $E(\,)$ operator can be intuitively interpreted as 'taking the average'). We will drop the subscript i in subsequent expression when defining the average effect. This could also be interpreted as the effect on the population by shifting the whole population from untreated to treated.

Another effect that is often relevant in applied research is the average treatment effect for the treated (ATT). This is the difference in the outcome that would have been observed if the individuals were treated and the outcome that would have been observed if they were not treated among those who were treated [i.e. $E\left(Y^1 - Y^0 | A = 1\right)$, where A is an indicator variable that A = 1 for the treated and A = 0 for the untreated].

In RCTs, ATT and ATE coincide because random allocation balances the treatment and control groups. With a sufficiently large sample, the only difference between groups will be in whether the participants received treatment. In non-randomized observational studies, by contrast, the participants often self-select themselves into treatment, and there are often systematic differences between those who took treatment and those who did not. Therefore, ATT and ATE differ in most non-randomized observational studies. The ATT is more relevant for interventions that are self-selected by individuals. For example, those who voluntarily engage in psychological interventions for alcohol use disorder are likely to have different characteristics compared to those who do not seek treatment. Therefore, it is more relevant to estimate the causal effect of such an intervention among those who were treated (i.e. ATT). However, for interventions that are designed to impact the whole population, such as increasing the price of alcohol, it is more relevant to estimate the causal effect on the population (i.e. ATE).

In RCTs, the ATE can be directly estimated from the average difference in the observed outcomes between the treatment/exposure and control conditions. In observational studies, treatment allocation and the outcome are often confounded. Applied researchers often 'adjust' or 'control' for confounding variables by adding these variables into a regression-based model. While, in theory, this approach can identify the casual effect of treatment, there are several practical limitations such as the linearity assumption not being met, and the adjustment relying on extrapolation. Also, with the outcome always in sight during the modelling process, researchers may be tempted to cherry-pick a model that produces the desired conclusion [6]. The methods discussed in the next sections address some of these limitations and thereby improve the validity of causal inference from non-randomized observational studies.

ASSUMPTION OF CAUSAL INFERENCE

Regardless of the methods used, the following assumptions are required to identify a causal effect [3, 7].

1. No interference—this assumes that whether an individual receives treatment does not affect the outcome of another individual. This assumption will be violated if there is spillover or a contagion effect.
2. Consistency—for each individual, only one of the two potential outcomes is observed. This assumption requires that the potential outcome under the observed treatment is the same as the observed outcome. This is a technical assumption needed to link the observed data to the potential outcomes framework.
3. No unmeasured confounding—this assumption requires that conditioning on a range of potential confounders treatment assignment is independent of the potential outcome, so that the treatment–outcome relationship is not confounded. This assumption is usually untestable and is often violated. The impact of its violation can be evaluated through sensitivity analyses [8]. For example, VanderWeele & Ding [8] introduced the E-value, which represents the minimum strength of association that an unmeasured confounder needs to have on both the treatment and outcome to explain away the treatment–outcome association. Therefore, a small E-value suggests that the presence of a weak unmeasured confounder will be sufficient to invalidate a causal interpretation a treatment–outcome association.
4. Positivity—this requires that treatment is not deterministic at every level of each of the covariates. As a result, individuals are assumed to always have some chance of receiving treatment, regardless of the values of covariates (e.g. age, gender and socio-economic status).

PART II: MATCHING

The goal of matching is to emulate the balance between treatment and control group in RCTs and ensure as much as possible that, after matching, the distributions of all observed covariates are similar in both treatment and control groups [9–11]. One-to-one matching is commonly used to estimate ATT. At its simplest, this method matches each individual in the treatment group to an individual in the control group based on a 'similarity' measure, such as propensity score [12, 13]. The propensity score can be intuitively considered as the probability of

receiving treatment calculated from each individual's observed covariates. It is often estimated using logistic regression with pre-treatment covariates. Two individuals, one from the treatment group and one from the control group, are considered to be a match if the difference in their propensity scores is smaller than a pre-determined threshold, known as the caliper (e.g. 0.02 [14]). Unmatched individuals are excluded from the analysis. As the treatment and the control are balanced after matching, a simple regression model with the outcome regressing only on treatment (treatment versus control) can estimate the ATT. It is recommended in this simple regression that cluster-robust standard errors be used for inference [15].

To illustrate the matching procedure, suppose we want to estimate the causal impact of smoking on psychological distress. The top panel in Table 1 shows the characteristics of people who smoked and people who did not from a simulated data set based on a national survey in Australia. Those who smoked and those who did not differed on several variables. For example, 42% of people who smoked versus 64% of people who did not smoke finished high school. To emulate the balance between those who smoked (treatment group) and those who did not (control group) that would have been achieved in a RCT, we matched each individual in the smoking group with an individual in the non-smoking group based on all the observed covariates (known as one-to-one matching). The matched sample demonstrates better balance between the two groups (e.g. 42 and 43% of the smoking and the non-smoking groups finished high school; see the right-most three columns in Table 1). Analysis can then proceed with the matched sample using linear regression and cluster-robust standard error for inference. For example, a linear regression shows that people who smoked report higher psychological distress than people who did not (see Supporting information, Appendix S1).

TABLE 1 Sample means in the unmatched and matched sample (top panel); sample means in the unweighted and weighted sample (bottom panel)

Matching						
	Unmatched sample mean			**Matched sample mean**		
	Smoke (*n* = 974)	**Not smoke (*n* = 7026)**	**Standardized mean difference**	**Smoke (*n* = 974)**	**Not smoke (*n* = 974)**	**Standardized mean difference**
Male	0.494	0.442	0.104	0.494	0.475	0.037
Indigenous status	0.052	0.018	0.157	0.052	0.048	0.018
Finished high school	0.422	0.638	−0.437	0.422	0.436	−0.029
Partnered	0.463	0.691	−0.458	0.463	0.444	0.039
Regionality						
Major cities	0.585	0.677	−0.187	0.585	0.587	−0.004
Inner regional	0.218	0.190	0.067	0.218	0.232	−0.035
Outer regional or more remote	0.197	0.133	0.162	0.197	0.181	0.041
English-speaking	0.958	0.913	0.223	0.958	0.954	0.021
Risky alcohol use	0.643	0.541	0.212	0.643	0.640	0.006
Age	51.606	53.782	−0.168	51.606	51.298	0.024
Inverse probability treatment weighting						
	Unweighted sample mean			**Weighted sample mean**		
	Smoke	**Not smoke**	**Standardized mean difference**	**Smoke**	**Not smoke**	**Standardized mean difference**
Male	0.494	0.442	0.104	0.452	0.449	0.007
Indigenous	0.052	0.018	0.239	0.020	0.021	−0.010
Finished high school	0.422	0.638	−0.443	0.607	0.612	−0.010
Partnered	0.463	0.691	−0.483	0.654	0.664	−0.020
Regionality						
Major cities	0.585	0.677	−0.195	0.664	0.666	−0.005
Inner regional	0.218	0.190	0.070	0.195	0.193	0.003
Outer regional or more remote	0.197	0.133	0.186	0.141	0.140	0.003
English-speaking	0.958	0.913	0.164	0.923	0.918	0.017
Risky alcohol use	0.643	0.541	0.204	0.566	0.554	0.024
Age	51.606	53.782	−0.153	53.399	53.519	−0.008

R codes and the simulated data set are provided in the on-line Supporting information.

The validity of a causal interpretation will hinge upon whether all key confounders (i.e. variables that have strong causal impact on both the exposure/uptake of treatment and the outcome) are measured and balanced between the treatment and control group, which is required by the no unmeasured confounding assumption. While this assumption is untestable, a researcher can calculate the E-value (a simple approach for sensitivity analysis) to evaluate how strong a confounding effect due to unmeasured confounding is required to invalidate a causal interpretation [8].

Simulation studies demonstrate that one-to-one matching using propensity scores can produce a good balance between treatment and control groups, and is sufficient for a range of scenarios [11]. Therefore, we focus upon propensity score-matching in the section. Other matching regimes include k-to-1 matching, exact matching and the use of Mahalanobis distance as similarity measures [11]. To estimate ATE, full matching can be used. Modern machine-learning algorithms, such as gradient boosting regression, can be used to calculate propensity scores [16]. For a review of the advantages and drawbacks of different matching methods and algorithms, readers can consult the reviews by Stuart [11] and Rosenbaum [17].

For one-to-one matching, a common concern among applied researchers is that a large number of unmatched observations in the control group may be discarded, leading to reduced power. However, the impact on power is usually small because power is largely driven by the smaller group (the treatment group in many scenarios), and retaining only the best matches could increase power because the treatment and control groups are more closely matched and similar after one-to-one matching [11].

Matching has several advantages over traditional regression analysis with covariate adjustment [6, 11]. First, it clearly separates the design and analysis phases. Calculating the propensity score can be considered as the 'design' phase of the study, like the randomization phase in RCTs. The outcome is not used in the matching procedure so researchers can fit multiple models to calculate the propensity scores and select the one that produces the best balance between treatment and control group. Only when a good balance is achieved does the researcher proceed to the analysis phase and compare outcomes between groups. As the outcome is not used during matching, this reduces the temptation to cherry-pick a model that produces the desired conclusion. Secondly, unlike regression models in which the 'adjusted effect' can be due to extrapolation, researchers directly compare each of the covariates between treatment and control group to ensure that a balance is achieved.

PART III: INVERSE PROBABILITY TREATMENT WEIGHT

Inverse probability treatment weight (IPTW) is a method that is closely related to the propensity score method [13, 18]. In the IPTW method, weighting is used to achieve a balance between the treatment and control groups. Weighted data can be thought of as a pseudo-population in which the only difference between groups is in whether they received treatment [19, 20]. For individuals in the treatment group, the weight is calculated as the inverse of their propensity score (1 divided by the propensity score). For individuals in the control group, the weight is the inverse of 1 minus their propensity score. These weights are called the unstabilized weight because when an individual who has a small probability of being in treatment ends up receiving treatment, the individual has a large weight, substantially inflating the variance of the causal effect estimates. Robins *et al.* [19] suggested obtaining a stabilized weight by multiplying the unstabilized weight by the unconditional probability of receiving treatment (the raw probability of receiving treatment in the sample, not conditional upon any covariates) for the individual in treatment and the unconditional probability of not receiving treatment for the individual in the control group. To further reduce the impact of large weights, the stabilized weight can be truncated to a less extreme value, such as the 5th or 95th percentile. This procedure could introduce a small bias in the final causal effect estimate, but effectively reduces variance [21]. Once the weight is calculated, the researcher can compare outcomes between treatment and control by regressing the outcome solely on the treatment allocation (treatment versus control) with weights. Robust standard error is required for correct causal inference.

We used the same example of smoking and psychological distress to illustrate IPTW. The bottom panel in Table 1 (right-most three columns) shows the characteristics of smokers and non-smokers in the pseudo-population created by weighting, which produced a much better balance between the groups. The causal effect of treatment can now be estimated with a weighted regression analysis using the pseudo-population generated using weighting. Example R codes are provided in Supporting information, Appendix S2.

Extension to multi-wave longitudinal analysis

The IPTW method can be easily extended to longitudinal analysis of data in which there could be time-varying confounding [19]. A time-varying confounder is influenced by previous exposure, and influences future exposure and outcome. For example, peer drinking is a time-varying confounder in a longitudinal study of parental alcohol supply during adolescence (exposure) on future alcohol use disorder in young adulthood (outcome). This is because peer drinking can (i) be influenced by previous parental supply, (ii) influence future parental supply and (iii) influence the risk of alcohol use disorder in young adulthood (Figure 1). Standard regression is not able to adjust for such time-varying confounding and can introduce bias into the estimates that could lead to contradictory conclusions [22]. Using the IPTW method, the overall causal effect of the exposure on the outcome can be estimated with a weighted regression analysis if we assign a weight to the individual in each wave and multiply these weights to form a final weight. An example estimating the impact of adolescent smoking on psychological distress in young adulthood is presented in Supporting information, Appendix S2 with example analysis code in R.

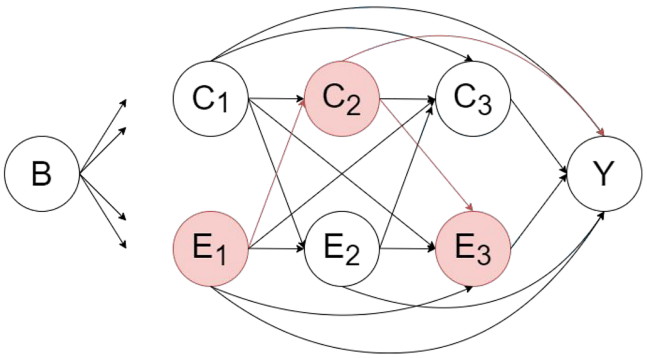


FIGURE 1 Association between the effect of parental alcohol supply during adolescence, peer alcohol use and future alcohol use disorder in young adulthood.B represents baseline demographic characteristics and covariates such as sex, baseline parental alcohol supply, baseline peer alcohol use, baseline risky alcohol use and baseline smoking status. C_1, C_2 and C_3 represent time-varying confounders including peer alcohol use and risky alcohol use at follow-up 1, 2 and 3. E_1, E_2 and E_3 represent the exposure, parental alcohol supply, at follow-up 1, 2 and 3. Y represents the outcome, which is alcohol use disorder in young adulthood. There is time-varying confounding because, for example, peer alcohol use and risky alcohol use at follow-up 2 (C_2) are influenced by parental alcohol supply at follow-up 1 (E_1), and they also confound the relationship between parental alcohol supply at follow-up 3 (E_3) and alcohol use disorder at young adulthood (Y)

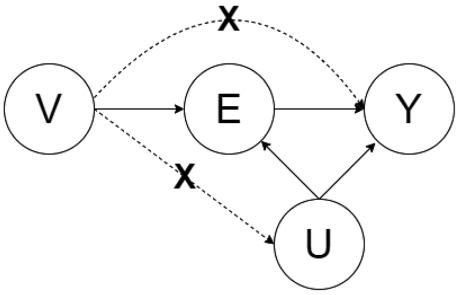


FIGURE 2 Association between an instrumental variable (V), treatment (E), outcome (Y) and unmeasured confounders (U). When testing the causal effect of alcohol consumption (E) on cardiovascular risk (Y) there would be many confounding variables, such as education, socio-economic status and life-style factors. The genetic variant of aldehyde dehydrogenase 2 family member (ALDH2) can be an instrumental variable because it influences alcohol consumption through a well-known biological mechanism. The (lack of) ALDH2 enzyme is unlikely to contribute to cardiovascular disease directly (hence the top dotted arrow). It is also unlikely to influence confounding variables (e.g. socio-economic status) that impact both alcohol use and cardiovascular disease (the dotted arrow from V to U)

PART IV: INSTRUMENTAL VARIABLE METHOD

Both matching and IPTW methods can be used to control for confounding from measured covariates but assume that there is no unmeasured confounding. This assumption is untestable and is likely to be violated because it is almost impossible to measure and include all possible confounding variables in the matching and weighting procedure. The instrumental variable method has been developed to control for unmeasured confounding [23, 24].

An instrumental variable is one that has (i) a direct causal impact on the exposure, (ii) no direct causal impact on the outcome and (iii) does not affect the outcome through variables other than the exposure (i.e. independently of any unmeasured confounders). The instrumental variable can therefore only affect the outcome through the exposure variable (Figure 2). This property is also known as the exclusion restriction assumption. Once a valid instrumental variable is identified, the researcher can first use a simple regression analysis to extract the variation in the treatment that is free of unmeasured confounding. Another regression then uses this confounding-free variation in the treatment to estimate the causal effect of treatment on the outcome.

To identify the causal effect using the instrumental variable method, the third assumption of no unmeasured confounding described in Part I needs to be adapted. It is now required that conditioning on a range of potential confounders, the instrumental variable is independent of the potential exposure so that the instrumental variable–exposure relationship is not confounded. Further, conditioning on a range of potential confounders and the potential exposure, the instrumental variable is independent of the potential outcome so that the instrument variable–outcome relationship is not confounded.

An example of instrumental variable in addiction research would be the use of genetic variants associated with alcohol use to estimate the effect of alcohol use on cardiovascular disease (see the footnote below Figure 2). Testing the causal impact of alcohol and cardiovascular disease is challenging, because it is unethical to randomize individuals to drink alcohol. Further, the association would be confounded by a large range of life-style, social and biological factors in observational studies. The aldehyde dehydrogenase 2 family member (ALDH2) gene can be used as an instrumental variable, because the ALDH2 enzyme is responsible for metabolizing alcohol by degrading acetaldehyde to nontoxic acetate. A variant of this gene (rs671 or ALDH2*2) results in an inactive product protein, impairing alcohol metabolism and resulting in an adverse reaction after drinking alcohol. This variant is therefore protective against alcohol use [25]. The ALDH2 gene is not associated with cardiovascular disease and is unlikely to be associated with confounding factors such as physical activity and socio-economic status. Therefore, if this gene were to affect cardiovascular outcomes, it must exert its effect solely through alcohol consumption. A simple comparison of cardiovascular disease between individuals with and without copies of the ALDH2 variant can test if alcohol consumption affects the risks of cardiovascular diseases. Further, by estimating the variation in alcohol consumption that is explained by the ALDH2 gene, researchers can also estimate the effect size of alcohol consumption on cardiovascular disease. This can be conducted using the two-stage least-square estimation (2SLS) with the following steps.

1. Regress the treatment (alcohol consumption) on the instrumental variable (ALDH2 gene) by least squares, and calculate the predicted value of treatment for each individual.
2. Regress the outcome (cardiovascular disease) on the predicted value of treatment (predicted value of alcohol consumption from step 1).

The coefficient of the predicted value of exposure represents the causal effect on the outcome. For inference, the standard error needs to be adjusted for the uncertainty in the predicted value of treatment. This adjustment has been implemented in several open-source software packages, such as the *ivpack* package in R [26]. Baiocchi *et al.* [24] provides an in-depth tutorial on the instrumental variable method with a detailed example in R using the *ivpack* package.

One challenge of the instrumental variable method is the identification of an instrumental variable. In the above example, the use of a genetic variant is a special application of the instrumental variable method, generally referred to as Mendelian randomization [27]. Individuals were effectively randomized at birth to have or not have a variant of the ALDH2 genes, which affected their alcohol consumption through a well-known biological mechanism [28]. As the ALDH2 genes affect the production of the ALDH2 enzyme, which is essential in alcohol metabolism, we can assume that it affects cardiovascular outcomes solely through alcohol consumption. However, the use of instrumental variables in other scenarios is challenging, because the exclusion restriction assumption is untestable and relies upon theoretical plausibility [24]. The assumption of no unmeasured confounding in conventional analysis such as regression and the above two causal analysis methods (matching and IPTW) is also untestable and often untenable. It has been argued that replacing the no unmeasured confounding assumption in conventional analysis with a theoretically justifiable assumption in instrumental variable analysis may provide stronger evidence for causal inference. Another example of a viable instrumental variable in addiction research is alcohol outlet density around an individual's residence. Higher alcohol outlet density is likely to encourage residents in the neighborhood to increase their alcohol consumption, and it is likely to be exclusively linked to health outcomes through alcohol consumption, after adjusting for socio-demographic factors.

There is still statistical consideration even if an instrumental variable satisfies all the necessary assumptions. If an instrumental variable is only weakly linked to the treatment (i.e. it can only explain a very small proportion of variance in the treatment variable), it will result in considerable variance in the estimates. Therefore, it is necessary to test the strength of the association between the instrumental variable and the treatment variable before conducting an instrumental variable analysis.

PART V: INTERRUPTED TIME-SERIES ANALYSIS

Many public health interventions are implemented to change a population-level outcome such as the rate of hospital emergency presentations due to excessive alcohol drinking. For example, O'Brien *et al.* [29] tested the effect of minimum alcohol pricing on population-level alcohol consumption in the Northern Territory, Australia.

Using data from repeated observations of an outcome, an interrupted time-series analysis (ITSA) compared the trend in the outcome before and after the intervention [30, 31]. This can be conceptualized based on the counterfactual framework as comparing what would have happened in the absence of the intervention (a counterfactual scenario) with what has been observed after an intervention. A control series, in which the intervention is not implemented, can be included to strengthen causal inference. The counterfactual scenario can be estimated based on the underlying trend preceding the intervention and the trend in the control series (including the time-points following intervention). A change in trend in the intervention series before and after the intervention, and a lack of change in the control series, provides good evidence for a causal effect of the intervention (Figure 3). Interrupted time-series analysis can control for secular trends in the outcome, confounding due to within- and between-group variation and fluctuation in the outcome due to seasonality. Estimates from interrupted time-series analysis are often comparable to estimates from RCTs [32, 33]. Further, it could have more important implications, because the data were from a naturalistic setting and so could potentially be more applicable in the real world.

The successful application of the interrupted time-series analysis requires a clear implementation time-point so that the pre-intervention period can be clearly differentiated from the post-intervention period. Further, it requires short-term outcomes that are responsive to the intervention. For example, in the evaluation of the effectiveness of a minimum alcohol pricing policy, the intervention time was clear because it was implemented on a specific date. For some designs, caution is required to allow for phase-in period that is clinically and/or theoretically derived. For example, using the current example individuals might stockpile alcohol when anticipating a price increase due to policy change.

Total alcohol consumption and alcohol sales data are ideal outcomes because they are usually very responsive to change in alcohol price and they are proximal to the policy change. Acute emergency department admissions may require a longer lead in period. Long-term outcomes such as changes in rates of liver cirrhosis are less suitable because they are more distal to the intervention.

It is also important that a comparable control series is chosen from a population that is as similar as possible to the one exposed to the intervention, except for the fact that one receives the intervention and one does not. For example, the control series can be from another location with similar population characteristics, or it can be from the historical trend in same population (e.g. comparing the time-series in the 18-month window where the intervention was implemented mid-way and the previous 18-month window when there was no intervention).

An interrupted time-series analysis with control is often based on the regression model:

$$Y_t = \beta_0 + \beta_1 \text{time} + \beta_2 \text{post intervention} + \beta_3 \text{post intervention} \times \text{time} + \beta_4 \text{group} + \beta_5 \text{group} \times \text{time} + \beta_6 \text{group} \times \text{post intervention} + \beta_7 \text{time} \times \text{post intervention} \times \text{group} + \epsilon_t$$

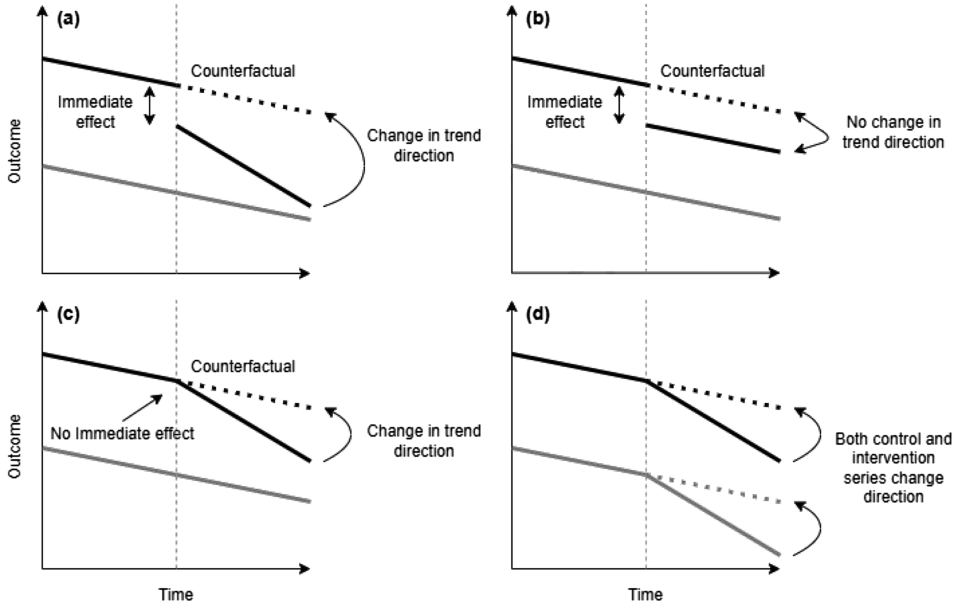


FIGURE 3 The black and grey line represents observed time-series from intervention and control group. In (a), there is an immediate effect of intervention, and there is a change in trend direction. The control series continues its original trajectory, providing strong evidence of a causal intervention effect. Similarly, in (b), there is an immediate effect but no change in trend direction; in (c) there is a change in trend direction but no immediate effect. In (d), although there is a change in trend direction after the intervention, the same change is also observed in the control series, indicating that the change is probably caused by other confounding factors, such as another intervention, that is implemented in both the intervention and control group

Post intervention is an indicator variable for pre- and post-intervention (0: pre-intervention; 1: post-intervention); group is an indicator variable for intervention and control group (0: control group; 1: intervention group); time is a numerical variable that represents the elapse of time. The interpretation of the coefficients β_0 to β_7 are explained in Figure 4. In this model, seasonality can be also adjusted for if indicator variables representing different seasons are added to model. For time-series, the outcome is often correlated over time and such autocorrelation must be accounted for. Several methods can be used. For example, the coefficients and the corresponding standard errors can be estimated using generalized least squares with a specified residual structure (e.g. AR [1]). The coefficients can also be estimated using ordinary least squares with heteroskedasticity auto-correlation consistent (HAC) standard error [34]. An example estimating the effect of minimum alcohol pricing on population-level alcohol consumption is presented in Supporting information, Appendix S4 with R codes.

In this section we have only focused upon a regression-based technique with which most addiction researchers were familiar. For applications of other time-series analysis technique, such as ARIMA and ARIMAX, the readers can consult the review by Beard *et al.* [35].

DISCUSSION

RCTs are the gold standard for establishing causality, but they are not always a feasible approach due to ethical, cost or logistical reasons. We provided an introduction, with simulated data sets and R codes, to four statistical methods that could help addiction researchers draw causal inferences based on observational data. However, it should be noted that statistical method is not a replacement for overall research design. The first two methods, matching and IPTW, are generally applied to individual-level data, and could be prone to bias due to unmeasured confounding. To reduce the impact of confounding in these designs, it is essential to identify as many potential confounders as possible based on established theoretical frameworks, and collect as much information as possible on these confounding variables for statistical adjustment (e.g. through matching and inverse probability treatment weight). Directed acyclic graphs (DAGs) have been increasingly used for confounder selection, but this requires adequate knowledge with regard to the underlying causal structure between variables [3]. However, this knowledge is often not available. VanderWeele [36] proposed that disjunctive cause criteria to simplify the confounder selection process. These criteria suggest adjusting for (1) all variables that were either a cause of the exposure or outcome or both, and excluding from this set any instrumental variable for the exposure and outcome, and (2) variables that were proxy for an unmeasured confounder that is a common cause for both the exposure and outcome.

Further, the Bradford–Hill criteria provide an overarching conceptual framework for evaluating overall evidence of the causal relationship between an exposure and an outcome in epidemiological research. Two of the criteria, biological plausibility and temporality, are particularly relevant for accessing causality in individual study. Biological plausibility refers to the presence of evidence for a possible biological mechanism linking the exposure to the outcome (e.g. smoking releases carcinogens, thus it is possible that it could lead

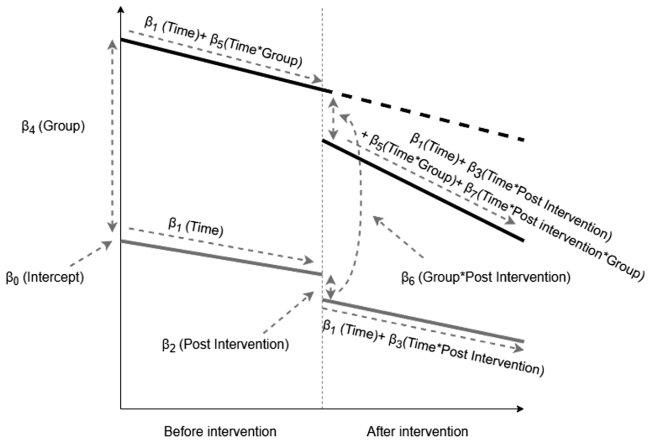


FIGURE 4 Coefficient interpretation from an interrupted time-series model. β_0 is the intercept of the model, which represents the outcome of the control group at time 0. β_1 is the slope of the control group before the implementation of the intervention in the control group. β_2 is the immediate change in the outcome in the control group after the intervention (this should be close to zero). β_3 is the change in slope in the control series after the intervention (this should be close to zero). Therefore, the slope in the control group after the intervention is $\beta_1 + \beta_3$. β_4 is the difference in the outcome at time 0 between the intervention and control group. Therefore, the outcome of the intervention group at time 0 is $\beta_0 + \beta_4$. β_5 is the difference in slope between the intervention and control group before the implementation of the intervention. Therefore, the slope of the intervention group before the intervention is $\beta_1 + \beta_5$. β_6 is the difference in the immediate change between the control and intervention group after the implementation of the intervention. β_7 is difference in the post-intervention slope between the intervention and control group compared to pre-intervention. Because the post-intervention slope of the control group is $\beta_1 + \beta_3$ and the pre-intervention slope of the intervention group is $\beta_1 + \beta_5$, the post-intervention slope of the intervention group is $\beta_1 + \beta_3 + \beta_5 + \beta_7$

to cancer). Temporality requires that the exposure preceding the outcome, and this can be addressed with careful longitudinal design. For example, three waves of data can be used for matching to strengthen causal inference—the exposure (measured at wave 2) can be matched based on a range of pre-exposure variables (measured at wave 1), and the exposure can be tested as a predictor for the outcome measured at a later time-point (e.g. wave 3). The third method, instrumental variable, is a powerful method that can overcome the issue caused by unmeasured confounding, but finding a suitable instrument could be challenging. The last method, interrupted time-series analysis, is generally applied to aggregated population-level data to evaluate policy impact. It is less prone to confounding at individual level and selection bias. However, it could be insensitive to detect the effect of an intervention at individual level and mask important differential effect on subpopulations.

Researchers should not blindly apply the methods discussed in this paper and always evaluate the plausibility of the assumptions underlying each method (e.g. assumptions about the relationship between the instrumental variable, exposure and outcome). These assumptions are often untestable using observational data and they need to be justified by plausibility. Sensitivity tests can help researchers to evaluate how serious the degree of assumption violation needs to be to invalidate their conclusions. Researchers should always consider sensitivity analysis in addition to their main analysis. To strengthen causal inference using observational studies, Hernán [37] outlined the target trial protocol, in which he proposed to apply thinking in RCT when designing an observational study, such as clearly specifying a priori the eligibility criteria, treatment strategies, outcome, causal estimates and statistical analysis.

CONCLUSION

Observational data can provide important information about causality. The four methods discussed in this paper can be used to strengthen causal inference from observational data, provided that the assumptions of each method are carefully considered and justified.

ACKNOWLEDGEMENTS

This study was supported by the Department of Health, Australian Government—National Health and Medical Research Council APP1176137. Open access publishing facilitated by The University of Queensland, as part of the Wiley - The University of Queensland agreement via the Council of Australian University Librarians.

DECLARATION OF INTERESTS

None.

AUTHOR CONTRIBUTION

Gary C. K. Chan: Conceptualisation: lead, methodology: lead, and writing - original draft: lead. **Carmen Lim:** writing - review & editing: supporting. **Tianze Sun:** writing - review & editing: supporting. **Daniel Stjepanovic:** writing - review & editing: supporting. **Jason Connor:** writing - review & editing: supporting. **Wayne Hall:** writing - review & editing: supporting. **Janni Leung:** conceptualisation: equal, investigation: equal, methodology: equal, and writing - original draft: supporting.

ORCID

Gary C. K. Chan https://orcid.org/0000-0002-7569-1948
Carmen Lim https://orcid.org/0000-0003-1595-6307
Tianze Sun https://orcid.org/0000-0002-3939-9499
Daniel Stjepanovic https://orcid.org/0000-0003-4307-423X
Jason Connor https://orcid.org/0000-0002-7020-1196
Wayne Hall https://orcid.org/0000-0003-1984-0096
Janni Leung https://orcid.org/0000-0001-5816-2959

REFERENCES

1. Goodman SN, Schneeweiss S, Baiocchi M. Using design thinking to differentiate useful from misleading evidence in observational research. Jama. 2017;317:705–7.
2. US Food and Drug Administration (FDA). Real-world Evidence Washington, DC: US FDA; 2021.
3. Pearl J. Causality Cambridge, UK: Cambridge University Press; 2009.

4. Rubin DB. Causal inference using potential outcomes: design, modeling, decisions. J Am Stat Assoc. 2005;100:322–31.
5. Winship C, Morgan SL. The estimation of causal effects from observational data. Annu Rev Sociol. 1999;25:59–706.
6. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivar Behav Res. 2011;46:399–424.
7. Hernán MA, Robins JM. Causal Inference Boca Raton, FL: CRC Press; 2010.
8. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. Ann Intern Med. 2017;167:268–74.
9. Rubin DB. Matched Sampling for Causal Effects Cambridge, UK: Cambridge University Press; 2006.
10. Rosenbaum PR, Rosenbaum P, Briskman. Design of Observational Studies New York, NY: Springer; 2010.
11. Stuart EA. Matching methods for causal inference: a review and a look forward. Stat Sci. 2010;25:1–21.
12. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. J Am Stat Assoc. 1984;79:516–24.
13. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70:41–55.
14. Austin PC. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. Pharm Stat. 2011;10:150–61.
15. Stuart EA, King G, Imai K, Ho D. MatchIt: nonparametric preprocessing for parametric causal inference. J Stat Softw. 2011;42:1–28.
16. Ridgeway G, McCaffrey D, Morral A, Burgette L, Griffin BA. Toolkit for Weighting and Analysis of Nonequivalent Groups: A tutorial for the twang Package Santa Monica, CA: RAND Corporation; 2017.
17. Rosenbaum PR. Modern algorithms for matching in observational studies. Annu Rev Stat Appl. 2020;7:43–176.
18. Rosenbaum PR. Model-based direct adjustment. J Am Stat Assoc. 1987;82:387–94.
19. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. 2000;11:550–60.
20. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Stat Med. 2015;34:3661–79.
21. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. Am J Epidemiol. 2008;168:656–64.
22. Suarez D, Borràs R, Basagaña X. Differences between marginal structural models and conventional models in their exposure effect estimates: a systematic review. Epidemiology. 2011;22:586–8.
23. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. J Am Stat Assoc. 1996;91:444–55.
24. Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. Stat Med. 2014;33:2297–340.
25. Connor JP, Haber PS, Hall WD. Alcohol use disorders. Lancet. 2016;87:988–98.
26. Jiang Y, Small D. *ivpack*: Instrumental Variable Estimation. The Comprehensive R Archive Network; 2021. Available from: https://cran.r-project.org/web/packages/ivpack/index.html. Accessed 14 June 2022.
27. Davies NM, Holmes MV, Smith GD. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ. 2018;362:k601.
28. Higuchi S, Matsushita S, Murayama M, Takagi S, Hayashida M. Alcohol and aldehyde dehydrogenase polymorphisms and the risk for alcoholism. Am J Psychiatry. 1995;152:1219–21.
29. O'Brien J, Tscharke B, Bade R, Chan G, Gerber C, Mueller J, et al. A wastewater-based assessment of the impact of a minimum unit price on population alcohol consumption in the Northern Territory. Addiction. 2022;117:243–9.
30. Bernal JL, Cummins S, Gasparrini A. The use of controls in interrupted time series studies of public health interventions. Int J Epidemiol. 2018;47:2082–93.
31. Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. Int J Epidemiol. 2017;46:348–55.
32. Fretheim A, Soumerai SB, Zhang F, Oxman AD, Ross-Degnan D. Interrupted time-series analysis yielded an effect estimate concordant with the cluster-randomized controlled trial result. J Clin Epidemiol. 2013;66:883–7.
33. St. Clair T, Cook TD, Hallberg K. Examining the internal validity and statistical precision of the comparative interrupted time series design by comparison with a randomized experiment. Am J Eval. 2014;35:311–27.
34. Andrews DW. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica. 1991;59:817–58.
35. Beard E, Marsden J, Brown J, Tombor I, Stapleton J, Michie S, et al. Understanding and using time series analyses in addiction research. Addiction. 2019;114:1866–84.
36. VanderWeele TJ. Principles of confounder selection. Eur J Epidemiol. 2019;34:211–9.
37. Hernán MA. Methods of public health research—strengthening causal inference from observational data. N Engl J Med. 2021;385:1345–8.

SUPPORTING INFORMATION

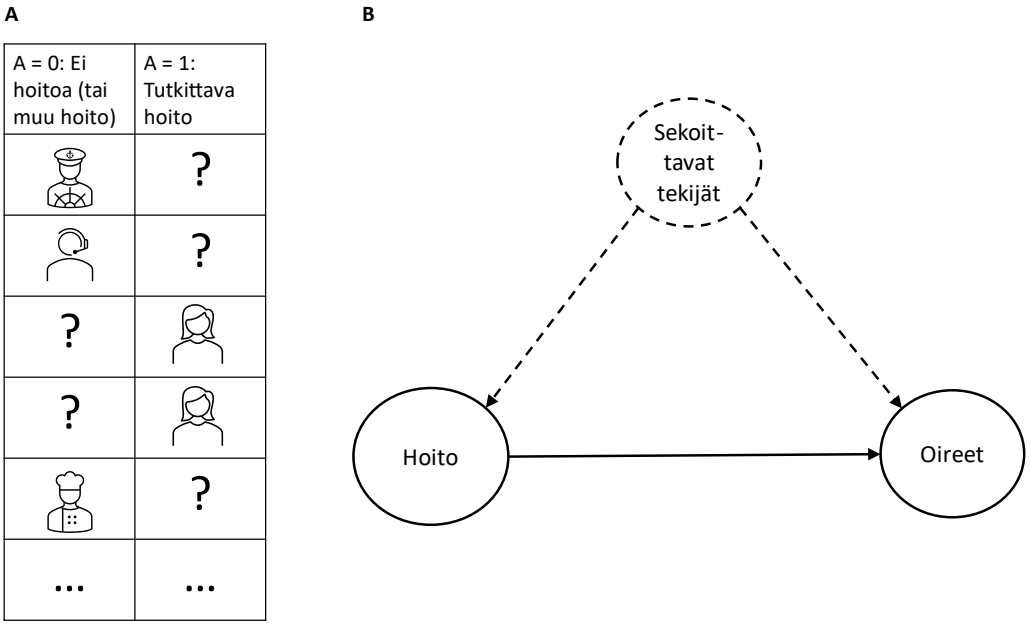
Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Chan GCK, Lim C, Sun T, Stjepanovic D, Connor J, Hall W, et al. Causal inference with observational data in addiction research. Addiction. 2022;117(10):2736–44. https://doi.org/10.1111/add.15972

Syy-seuraussuhteiden (kausaliteetin) päättely

Vaikka toisinaan onkin hyödyllistä tuntea jonkin asian ennusmerkkejä, selittävässä tieteessä ollaan useimmiten kiinnostuneempia asian syytekijöistä. Esimerkiksi, jos tiedämme henkilön käyneen masennuksen psykoterapiahoidon, hänen masennusoireistonsa on todennäköisesti edelleen koholla suhteessa satunnaisesti väestöstä valittuun henkilöön. Hoito ei kuitenkaan ollut toimimaton. Vertailuryhmä vain on väärä. Psykoterapiaan päätyvät ihmiset keskimäärin eroavat paljon satunnaisesti valitusta väestön edustajasta. Heidän oireensa voivat merkittävästi laskea hoidon aikana ja silti jäädä oireetonta korkeammalle tasolle. Vakuutusmatemaatikon näkökulmasta hoito voi olla hyvä oireilun ja sairastelun riski-indikaattori, eli ennusmerkki. Hoidon tuottajan näkökulmasta syy-seurausvaikutus on olennaisempi: kuinka potilaalla olisi mennyt, jos juuri hän ei olisi saanut hoitoa? Toisin sanoen, onko hoito auttanut henkilöä?

Syy-seurausvaikutusta voidaan havainnollistaa ideaalisen kokeen avulla (Kuva 1A). Niin kokeellisissa kuin havaintoaineistoissakin, näemme jokaiselle yksikölle (esim. henkilölle) jonkin *lopputuleman* (*engl.* outcome; esim. oireet) hänen saamalleen *altistukselle* (*engl.* exposure; esim. hoito). Emme kuitenkaan näe mitä juuri kyseinen henkilö olisi saanut toisen *potentiaalisen* altistuksen tapauksessa. Kuitenkin juuri ero tuohon *faktanvastaiseen* (kontrafaktuaaliseen) tapahtumaan olennaisesti määrittää *vaikuttiko* altistus lopputulemaan. Esimerkiksi psykoterapiaan päätyvistä noin 70 % on naisia ja naisilla on myös noin kaksinkertainen masennusriski miehiin nähden. Yleisväestössä naisia on vain noin 50 %. Sukupuoli saattaa siten vaikuttaa sekä hoitoon päätymiseen että tutkittavaan lopputulemaan. Se on siis mahdollinen *sekoittava tekijä* (*engl.* confounder, confounding factor).



Kuva 1. Syy-seurauspäättely. **A**) *Ideaalinen koe.* **B**) *Suunnattu graafi.*

Tutkimuksessa syy-seurausvaikutuksia pyritään usein päättelemään vertailemalla altistuksen saaneen ryhmän lopputulemaa ryhmään, jota ei altistettu. Esimerkiksi psykoterapian saaneiden oireita verrataan niiden oireisiin, jotka eivät saaneet psykoterapiaa. ”Sekoittava tekijä” sekoittaa tämän vertailun tuottaman syy-seuraustiedon. Esimerkiksi verratessamme psykoterapiapotilaiden ryhmää, jossa naiset ovat yliedustettuna yleisväestön ei-potilaisiin, tahtomattamme vertaamme myös naisia miehiin. Tämän seurauksena emme voi eristää hoidon vaikutusta lopputulemaan sitä sekoittavasta

sukupuolen vaikutuksesta. Sukupuolen vaikuttaessa sekä altistustodennäköisyyteen että lopputulemaan, saatamme virheellisesti päätellä hoidon toimivuuden heikoksi siksi, että hoitoryhmässä on lähtökohtaisesti oireisempia yksilöitä. Muuttujien välisiä riippuvuussuhteita havainnollistetaan usein ns. *suunnatulla graafilla* (Kuva 1B), jossa pallot kuvaavat muuttujia ja nuolet teoreettisia syy-seuraussuhteita, eli sitä mikä muuttuja vaikuttaa mihinkin. Erityisen ongelmallista tutkimukselle on, ettei kaikkia sekoittavia tekijöitä välttämättä havaita, tai edes tunneta (katkoviivat kuvassa).

Kuvasta 1A nähdään, että hoitoryhmän keskiarvo heijastelee keskimäärin enemmän koulutettujen naisten tilannetta, kun taas muussa väestössä myös ammattimiehet ovat paremmin edustettuina. Erilaisia tunnettuja ja tuntemattomia sekoittavia tekijöitä on siis oltava. *Satunnaistetun kokeen* keskeinen idea on tasapainottaa altistusryhmät arpomalla, jolloin tiedämme, ettei mikään syytekijä voi systemaattisesti aiheuttaa sekä altistuksen todennäköisyyttä että lopputulemaa (eli olla sekoittava tekijä). Esimerkiksi, otamme psykoterapiaan halukkaita ja ohjaamme heitä arpomalla kahteen ryhmään, hoidon saaviin (hoitoryhmä; *engl*. treatment group) ja hoidolta evättyihin (kontrolliryhmä; *engl*. control group). Kontrolliryhmä voi olla myös vaihtoehtoista hoitoa saavien ryhmä. Keskeistä on, ettei mikään syytekijä ole voinut systemaattisesti vaikuttaa ryhmien jäsenrakenteeseen. Tällöin, riittävän suuressa otoksessa, ryhmät edustavat tasapuolisesti muita syytekijöitä ja eroavat systemaattisesti vain tutkitun syytekijän osalta. Ryhmien keskimääräinen ero lopputulemassa siis mittaa keskimääräistä hoitovaikutusta (*engl*. average treatment effect, eli ATE).

Monessa tapauksessa satunnaistetun kokeen järjestäminen ja kontrollointi on lähes mahdotonta tai tarpeettoman kallista. Onneksi on olemassa myös vaihtoehtoisia tapoja arvioida keskimääräisiä hoitovaikutuksia ja muita eksoottisempiakin syy-seurausvaikutuksia. Chan ym. (2022) tutoriaaliartikkeli käy läpi joitain näistä vaihtoehdoista. Alla on artikkelin tarkempi viite sekä joidenkin sen avainkäsitteiden käännöksiä.

Lähteet

Chan, G. C. K., Lim, C., Sun, T., Stjepanovic, D., Connor, J., Hall, W., & Leung, J. (2022). Causal inference with observational data in addiction research. *Addiction*, 117(10), 2736–2744. https://doi.org/10.1111/add.15972

Käännökset

- Association
    - Yhteys
    - Association
- Assumption
    - Oletus
    - Antagande
- Average treatment effect for the treated
    - Keskimääräinen käsittelyvaikutus altistuneilla, keskimääräinen hoitovaikutus hoidetuilla
    - Genomsnittlig behandlingseffekt för de exponerade, genomsnittlig behandlingseffekt för de behandlade

- Causal
    - Syy-seuraus, kausaalinen
    - Kausal, som gäller orsak och verkan
- Causal inference
    - Syy-seurauspäättely, kausaalipäättely
    - Kausal inferens, kausal slutledning
- Cause
    - Syy, aiheuttaa
    - Orsak, orsaka
- Coefficient (a constant by which an algebraic term is multiplied; e.g., regression coefficient)
    - Kerroin (vakio, jolla algebrallista termiä kerrotaan; esim. regressiokerroin)
    - Koefficient (en konstant som en algebraisk term multipliceras med; t.ex. regressionskoefficient)
- Confounding factor, confounding variable
    - Sekoittava tekijä/muuttuja
    - Förväxlingsfaktor, -variabel
- Consistency
    - Johdonmukaisuus
    - Följdriktighet
- Control group
    - Kontrolliryhmä
    - Kontrollgrupp
- Counterfactual
    - Kontrafaktuaalinen, faktanvastainen, tosiasioiden vastainen
    - Kontrafaktisk, faktamotsägande
- Covariate
    - Kovariaatti, selittävä muuttuja
    - Kovariat, förklarande variabel
- Directed acyclic graph
    - Suunnattu ei-syklinen graafi
    - Riktad acyklisk graf
- Exposure
    - Altistus, altistemuuttuja
    - Exponering, exponeringsvariabel
- Instrumental variable method
    - Välinemuuttuja-analyysi
    - Instrumentvariabelmetod
- Interference
    - Interferenssi, väliintulo
    - Interferens
- Interrupted time-series analysis
    - Keskeytetyn aikasarjan analyysi
    - Analysmetoden avbruten tidsserie
- Intervention
    - Interventio
    - Intervention
- Inverse probability

    - Käänteistodennäköisyys
    - Omvänd sannolikhet
- Inverse probability of treatment weight
    - Käänteistodennäköisyyspainotus (hoidon suhteen)
    - Omvänd sannolikhetsviktning (i förhållande till behandling)
- Longitudinal
    - Pitkittäinen, pitkittäis-
    - Longitudinell
- Matching
    - Kaltaistus
    - Matchning
- Potential outcome
    - Mahdollinen (potentiaalinen) lopputulema
    - Potentiellt utfall
- Propensity score
    - Taipumuspistemäärä
    - Benägenhetspoäng
- Randomized controlled trial (RCT)
    - Satunnaistettu vertailukoe
    - Randomiserad kontrollerad prövning
- Selection bias
    - Valikoitumisharha, valintaharha
    - Urvalsbias
- Stabilized weight
    - Vakautettu painokerroin
    - Stabiliserad viktkoefficient
- Standardized mean difference
    - Standardoitu keskiarvoero
    - Standardiserad genomsnittlig skillnad
- Treatment
    - Hoito
    - Behandling
- Truncate
    - Katkaista
    - Trunkera
- Weighting
    - Painotus, painottaminen
    - Viktning