Cross-Sectional Model-Building for Research on Subjective Well-Being: Gaining Clarity on Control Variables

David Bartram[1]

Abstract

Happiness/well-being researchers who use quantitative analysis often do not give persuasive reasons why particular variables should be included as controls in their cross-sectional models. One commonly sees notions of a "standard set" of controls, or the "usual suspects", etc. These notions are not coherent and can lead to results that are significantly biased with respect to a genuine causal relationship.

This article presents some core principles for making more effective decisions of that sort. The contribution is to introduce a framework (the "causal revolution", e.g. Pearl and Mackenzie 2018) unfamiliar to many social scientists (though well established in epidemiology) and to show how it can be put into practice for empirical analysis of causal questions. In simplified form, the core principles are: control for confounding variables, and do not control for intervening variables or colliders. A more comprehensive approach uses directed acyclic graphs (DAGs) to discern models that meet a minimum/efficient criterion for identification of causal effects.

The article demonstrates this mode of analysis via a stylized investigation of the effect of unemployment on happiness. Most researchers would include other determinants of happiness as controls for this purpose. One such determinant is income—but income is an intervening variable in the path from unemployment to happiness, and including it leads to substantial bias. Other commonly-used variables are simply unnecessary, e.g. religiosity and sex. From this perspective, identifying the effect of unemployment on happiness requires controlling only for age and education; a small (parsimonious) model is evidently preferable to a more complex one in this instance.

Keywords Happiness · Control variables · Causal relationships · Cross-sectional data

1 Introduction

This article is intended to help researchers make better decisions in building cross-sectional models intended to explore the causal impact of particular factors on subjective well-being (SWB, considered here as happiness). Currently, some researchers invoke the notion

---

✉ David Bartram
    d.bartram@le.ac.uk

[1] University of Leicester, Leicester, UK

of a "standard set"[1] of variables affecting happiness—and they proceed to include all of these variables as controls in their models. In some instances, researchers include controls without articulating any particular reason for doing so. Happiness researchers are far from alone in this respect: as Pearl and Mackenzie (2018: 139) note, researchers and statisticians have been "immensely confused" for an extended period about what to include as controls (see also Gangl 2010). Much greater clarity has emerged in recent years via a "causal revolution" (Pearl and Mackenzie 2018)—a perspective now well established in epidemiology (e.g. Arnold et al. 2020) but only slowly gaining traction in the social sciences (e.g. Glynn and Kashin 2017.)

The framework described here centres on the proposition that models exploring the causal effect of a particular variable should include (as controls) only factors that act as confounders with respect to that effect. For this purpose there is no standard set—and it might well be reasonable to omit variables even when in general they have significant effects on happiness. At a minimum, researchers should articulate reasons for including variables as controls—and the only sensible reason (in connection with identifying a causal relationship) is that a (potential) control is a confounder, such that omitting it would result in a biased estimate of the effect we are seeking to gauge.

A related goal is to help researchers avoid a specific error. Happiness researchers (in common with scholars working on other topics) sometimes construct models that include other variables that intervene in a path from the main variable to happiness (the outcome variable). If that path involves significant indirect effects (significant in the substantive sense, not just the statistical sense), the result is likely a biased estimate of the total effect—which is usually the quantity researchers are seeking to investigate. In situations of this sort, we must omit a variable even if it has substantial effects on the dependent variable (e.g. Smith 1990; Gangl 2010).

A more explicit engagement with the topic of causation and causal modelling will offer greater clarity and help researchers avoid errors of that latter sort. Usually that topic is presented via complex mathematical notation and formal logic (e.g. Gangl 2010; VanderWeele and Shpitser 2011); even "friendly" treatments (e.g. Pearl 2009) are sometimes unapproachable. This article avoids the usual apparatus entirely and proceeds mainly via a straightforward example. Instead of elaborate mathematical notation, this article use graphical causal diagrams (sometimes called directed acyclic graphs, or DAGs, e.g. Elwert 2013) to represent causal models and to gain clarity on the relationships among variables—in particular, whether a specific variable is a confounder, an intervening variable, or neither.

The arguments developed here are especially pertinent for analysis of cross-sectional data. When using panel data and longitudinal models, researchers are sometimes more attuned to the question of what variables to include, and why; change over time is a salient feature of the data, so we might find greater clarity on how to think about why changes occur and how to explain them properly. There are significant advantages to using panel data; where a genuine choice exists, it would be hard to justify using cross-sectional data instead. But for some questions exploring SWB, only cross-sectional data are available; this is certainly the case for most research involving international comparisons. The European

---

[1] The term "standard set" (referring to the way control variables were selected for inclusion) appears e.g. in the work of Hou (2014), Gonza and Burger (2017), and Devine et al. (2019). It is perhaps unfair to single out these authors in this context; others adopt the same approach but use different terms to describe it (I adopted it myself in earlier work: Bartram 2011). This mode of analysis is far from uncommon.

Social Survey and the World Values Survey are important resources for some happiness researchers—and it is therefore important to use them correctly.

The focus here is on single-stage regression models (as against two-stage techniques e.g. instrumental variables or treatment effects models). Many researchers believe that models of that sort cannot tell us anything about causality; one commonly sees disclaimers about causal interpretations. Under certain conditions, this view is unnecessarily cautious; depending on the assumptions we are willing to make, a regression model that controls only for confounders might provide a reasonable foundation for discerning a causal effect. Use of more sophisticated techniques is laudable, and quantitative researchers should of course aspire to improve their abilities. But we can anticipate that many researchers will continue to use single-stage models, and so it is worthwhile trying to improve practice in this regard as well. When researchers do not consider the relationships among the various independent variables in their analysis, the results likely tell us little about causal effects (so, their disclaimers are to be taken quite seriously). The approach described here will help researchers construct cross-sectional analyses offering results that need not be summarily dismissed.

The article works with a stylized exploration of the impact of unemployment on happiness. It does not offer new substantive insights into that relationship. We therefore forego the usual literature review – the contribution on offer does not pertain to unemployment and happiness per se but rather to the task of model building intended to gauge causal effects. So, we proceed in the next section with a consideration of how researchers frame their core research goals and what sort of analytical approach is entailed by those goals.

2 Causal Explanation: Ambitions and Methods

Happiness researchers commonly want to know what "determines" or "affects" SWB. Those words and others similar to them put us squarely in the realm of seeking causal explanations. That point is uncomfortable for some (especially some sociologists), but there is no satisfactory or desirable way around it. On this question social researchers face a choice: either the patterns apparent in social life are the result of an astonishing set of coincidences – or there are causal relationships that produce them (even if we do not yet fully understand those relationships). Some researchers try to avoid the point by using different terms: perhaps some factors merely "shape" a situation. But that word is nothing more than a euphemism and the core point remains.[2] Intuitively, our interest in causal relations is evident in the way we answer a question starting with "why"?: in everyday (English) language, we typically begin our answer with "be*cause*".

A causal relationship is usefully defined with reference to the notion of "counterfactual dependence": we can discern a causal effect of X on Y if we are confident that if X had been different Y would have been different as well (Freese and Kevern 2014). Researchers sometimes express their findings via the term association. That term is much more general, to such an extent that it often hides more than it reveals. An "association" is best considered a merely mathematical entity with limited potential for giving us insight about the social world. Finding an association between unemployment and happiness is not a good

---

[2] I am grateful to my colleague Patrick White for insightful crystallization of these observations.

stopping-point for research: we surely want to know, what *kind* of association? Does unemployment lead to happiness? Or vice-versa? (Or both?).

We may then consider: what sort of analysis (quantitative or otherwise) can count as evidence in support of a claim about a causal relationship? There is an enormous literature addressing that question (e.g. Lieberson 1985; Morgan and Winship 2007; Pearl and Mackenzie 2018), and for the most part I will not engage it here; instead the main goal is to consider a specific point that arises when we have decided to use quantitative analysis of observational data for this purpose (again that point, developed below, is: control only for confounders). In that circumstance, there are as noted good reasons to use longitudinal analysis of panel data, where it exists—but despite its advantages it offers no guarantees of estimates free from bias; even fixed-effects models can lead to biased estimates under certain conditions (Dieleman and Templin 2014; Bell et al. 2019). The question considered here is: what are the possibilities for using cross-sectional data for this purpose? Again, that question is important given the widespread use of such data in the field of happiness studies.

2.1 Using Cross-Sectional Data

Most researchers are aware of the potential pitfalls of using cross-sectional analysis to explore causal claims. In that context, we commonly see disclaimers about the direction of causality; unemployment might lead to lower happiness, but less-happy people might find it more difficult to get a job. Sometimes we see another, more sweeping disclaimer: the findings (e.g. from a regression model) are merely "correlations", with no assertion intended regarding causation. But statements of that sort simply evade the issue and undermine the value of one's work. In many instances, the motivation to investigate an association e.g. between unemployment and happiness is the possibility that one actually causes the other. There is again an intuition underpinning that assertion: researchers who offer disclaimers sometimes struggle to adhere strictly to those disclaimers in the rest of what they write, elsewhere using terms that strongly imply causality (as above: affects, determines, leads to, etc.). This tendency betrays the nature of their true interest. Our analyses should be constructed with a clear sense of the real purpose we have in mind for it. Sometimes one's goal is properly described as description (see Berk 2010), in which case the core arguments in this article might not apply. But in many such instances description is a prelude to explanation: we want to know, to the extent possible given the limitations of the data, how to explain the patterns apparent in our descriptions.

To gain insight into how this goal might be achieved, we can take as a starting point the bivariate relationship. Simplifying from a general observation by Pearl (2009): as an initial approximation, the effect of being unemployed is given by the difference in average happiness across those who are unemployed and those who are not. The (average) happiness of those who are employed stands as a naive counterfactual for the happiness of those who are unemployed; it indicates what the happiness of the unemployed might have been if they weren't unemployed (cf. Morgan and Winship 2007). The difference is then a first step in determining what the impact of *becoming* unemployed is—i.e., the (unobservable) change in one's happiness following from the (unobserved) change in one's employment status.

Drawing an initial connection between the two quantities is reasonable as a first step (even if we will then move quickly beyond it). If unemployed people are on average less happy than employed people, it would be unwise and unnecessary to conclude that this

raw difference has no value in learning about the impact of becoming unemployed. The raw difference is instead an insufficient basis for drawing firm conclusions—but we should be clear on *why* it is insufficient and what next steps are required. In particular, we should consider threats to causal inference associated with a particular approach to analysis (e.g. a simple/raw difference), and what sort of further analysis can be used to mitigate those threats.

The idea of threats is usefully specified as bias: a raw difference in happiness between the unemployed and the employed might suggest an effect of unemployment that is different in magnitude from the "true" effect—perhaps to such an extent that in reality there is no effect at all. The topic of bias is complex, with numerous typologies (e.g. Shahar and Shahar 2013). We focus here on "confounding bias", because this is the type most researchers intend to mitigate via use of regression models containing control variables. Confounding bias starts with the possibility that the observed correlation might be artificial, in the sense that there is some other variable (or set of variables) that determines unemployment as well as happiness; bias results from not taking account of that third factor. An example sometimes used for teaching starts with the observed correlation between height and math ability (taller people are in fact better at math): the association is readily understood as not indicative of a causal relationship, once we understand how age determines both (i.e., among children).

Regression models are used to try to ensure that any causal inference is reasonably robust to the possibility that the observed association is biased or even entirely spurious (with reference to a true causal effect). We adjust the raw differences via introduction of control variables, to reduce the likelihood of bias: controlling for age (comparing children at the same ages), we see that there is no causal relationship between height and math ability. Researchers who use regression models should have clarity about their purpose in these terms – as against proceeding from bivariate analysis to regression models only because a bivariate analysis would seem too simple to be credible, or proceeding to regression models by default, with no clear sense of why. So, again, the point of constructing a regression model including control variables is to mitigate against bias (confounding bias, in particular) and thus to arrive at a finding that is unbiased (or, more modestly, a less biased indication of a causal relationship than a bivariate/raw association would be).

2.2 Model Specification and Control Variables

Quantitative researchers understand these points in general terms. But a key difficulty sometimes arises in the application to particular analyses. It is evident that some researchers working on subjective well-being do not have a clear sense of how to effectively answer a key question: if regression models containing control variables are to be used for the purpose of mitigating bias, *which* control variables should be included?

That assertion is evident in the way some happiness researchers write their methods sections. All too often, we see a set of control variables identified as standard determinants of happiness. The list typically includes most or all of the following items: age, sex, partner status, education, employment status, income, religiosity, social connections, and perhaps a set of personality variables e.g. the "big five". Whatever one's research question (i.e., no matter what the main independent variable is), these are then used (almost by default) as the control variables. A regression model then gives a coefficient for the main independent variable that is interpreted as a "net effect" of that main variable—and this "net effect" is understood to be a genuine causal effect (in contrast to what appears in a bivariate model).
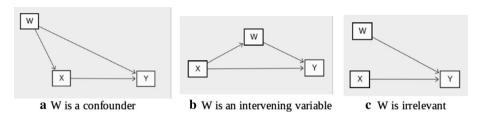
a W is a confounder   **b** W is an intervening variable   **c** W is irrelevant

Fig. 1 Simple Causal Diagrams. *Note*: in these DAGs, Y is the dependent variable, and the goal is to identify the effect of X on Y. W is a potential control variable with respect to that purpose

This way of seeing things implies that control variables work in just one way. This is a misunderstanding. Adding a control will have different implications for the result we see for the main independent variable, depending on the relationship between the control and that variable. The key question is: is the control a *determinant of* the main independent variable? Or, is the control *determined by* the main independent variable? Fig. 1 represents the relevant patterns; here, Y is the dependent variable, X is the main independent variable, and W is a potential control.

In 1.a, the control W functions properly to mitigate bias. The reason age makes sense as a control variable for discerning the (zero) effect of height on math ability is that age is a proper confounder: it is a determinant not only of the dependent variable (math ability) but also of the main independent variable (height). Age must be included here, to avoid bias.

In 1.b, we see a very different pattern, where the control intervenes in the relationship between the main independent variable and the dependent variable. An example would involve controlling for income to estimate the effect of unemployment on happiness: here, unemployment results in decreased income, which in turn results in decreased happiness. (The pattern would also suggest a direct effect of unemployment on happiness, separate from its indirect impact via income.) On this basis, it is incorrect to include income as a control variable; inclusion of an intervening variable will *induce* bias, not mitigate it. (This point is explored in more detail below). In 1.c, W is a cause of Y but there is no relationship between W and X. In this instance, inclusion of W as a control will not lead to biased estimates of the effect of X on Y—but as it is not a confounder with respect to that relationship it is unnecessary to include it (i.e., one's estimate of the impact of X on Y will not be biased via omission of W).

The key point missed by some quantitatively-oriented happiness researchers is that a set of control variables selected to reduce/mitigate bias must include only those variables that are understood to be antecedents of the main independent variable whose impact on happiness we are seeking to evaluate (Elwert & Winship 2014; Berk 1983). A variable that intervenes between the independent variable and happiness (or indeed any dependent variable) must not be included when estimating the overall impact of the main independent variable (e.g. Smith 1990; Gangl 2010; Wooldridge 2005). An intervening variable might be included at a later stage, when we try to gain clarity (e.g. with mediation analysis) on the mechanism(s) for a causal impact—but the impact itself must be determined via models that exclude intervening variables.

A more conventional treatment of these matters uses mathematical notation giving a formula corresponding to the regression model we implement in a statistical package. In Eq. (1), where X_1 is the main independent variable of interest, the other Xs (X_2, X_3, etc.) are control variables and ε is the error term (capturing residual variance in the outcome Y, i.e., variance not explained by the variables included in the model). A key assumption is

that the error term is uncorrelated with the independent variables in the model. In practice, that assumption amounts to a condition that must hold for unbiased estimation and effective interpretation: one must not have omitted determinants of the outcome that are correlated with variables in the model.

$$Y = a + b_1 X_1 + b_2 X_2 + b_3 X_3 + \cdots + b_k X_k + \varepsilon \tag{1}$$

This conventional notation is however insufficient and potentially misleading in an important way: it identifies control variables under the label of X without inquiring about the nature of the relationships between those Xs (the controls) and X_1, the main independent variable of interest. By considering all independent variables as so many Xs, it implies an equivalence in their status that is unlikely to hold in many situations. Introducing two controls (say, X_2 and X_3) can have quite distinct consequences for the interpretation of b_1 (the coefficient for X_1) if X_2 is a confounder and X_3 is an intervening variable.

For some quantitative researchers, a distinction between confounders and intervening variables will be deemed insufficient for research intended to evaluate causal relationships. With observational data, the criterion accepted by some as sufficient for causal identification is more stringent, labelled with terms such as "conditional independence" or "strict ignorability" (Gangl 2010). As noted, it is possible to offer more complex analyses (e.g. matching and instrumental variables models) that attempt to satisfy that criterion more robustly. Empirical contributions that succeed in this respect are of course valuable. The position offered here is both less ambitious and more achievable, with a focus on helping researchers avoid an important error that significantly impairs their engagement with more basic forms of cross-sectional analysis: when selecting control variables we must make a distinction between confounders and intervening variables.

3 Demonstration

Adoption of the guidelines above would lead to an improvement (perhaps even a substantial one) in research investigating the effect of factors conceived as potential determinants of happiness. To reiterate, confounding variables must be included as controls in regression models; not including them will lead to biased estimates of causal effects (sometimes called omitted variable bias). Inclusion of intervening variables, in contrast, will introduce/exacerbate bias (sometimes called "overcontrol bias", e.g. Rohrer 2018, or "included variable bias", York 2018). Variables that do not intervene but are also not confounders are unnecessary as controls. The analysis in this section illustrates these points via a stylized cross-sectional investigation of the relationship between unemployment and happiness. Again, the goal is not to add anything new to our understanding of that relationship (good research on the topic uses superior techniques applied to panel data e.g. the GSOEP). Instead the objective is to show the consequences of incorrect model-building particularly when using cross-sectional data—the sort of consequences that would emerge in research on other topics as well.

For this purpose I offer an analysis of data drawn from the British sample of the European Social Survey (ESS). Focusing only on a single country, we can avoid irrelevant complications that arise when multiple countries are included. I use the ESS's education variable—and a consistent format for the main education variable is available only as of Round 5. The data are therefore drawn from Rounds 5 through 8 (pertaining to the years 2010, 2012, 2014, and 2016; see Jowell 2007).

Happiness is measured via a question asking "Taking all things together, how happy would you say you are?", with answers drawn from an 11-point scale (zero for "extremely unhappy" to ten for "extremely happy"). Basic demographic variables include sex and age; an age-squared term is used as well. Additional variables include partnership status (constructed as binary: whether living with a partner or not), religiosity ("how religious would you say you are", using an 11-point scale similar to the happiness variable), and education (the standard EISCED categories). An income variable is constructed via nationally-derived deciles; UK respondents select from ten ranges given in British pounds.

Unemployment status is drawn from the "main activity" variable in the ESS. Respondents are "unemployed" in the analysis below whether they indicate they are actively looking for a job or not. The analysis compares unemployed people to those in paid employment (including self-employed). To simplify the presentation, I omit people who are not in the labour force (students, retired people, people working in the home, etc.).

The analysis consists of ordinary least-squares (OLS) regression models. (So, the mathematical expression of these models corresponds to Eq. 1 as given above.) As per Ferrer-i-Carbonell and Frijters (2004), ordered probit models would lead to the same substantive conclusions; a discussion of effect sizes is essential for our purposes, and OLS models are easier to interpret in that respect. The total sample size across all four rounds is 4045 (after deleting observations where there are missing values, so that the same sample is analysed across all models). All regression analyses use sampling (design) weights and include dummy variables for survey round.

The analysis begins with a regression model (1) in Table 1 that includes only (un)employment status and survey round; here unemployed people are less happy than those who are employed, by an average of 0.68 points on the 11-point scale.

If we were prepared to believe that the employed are a reasonable counterfactual (giving us a basis for knowing how happy the unemployed would be if they were not unemployed), we could conclude via Model 1 that unemployment leads to a decrease in happiness of 0.68 points (on the scale from 0 to 10). That is, if unemployed people were employed, the basic regression model suggests that they would on average report happiness at the average level of those respondents who are in fact employed. That counterfactual idea underpins a conclusion about the happiness impact of *becoming* unemployed – here, suggested as a decline of 0.68 points. But there is an obvious need to consider control variables before jumping to that conclusion; people who are unemployed might be different in other respects, and those other variables could contribute to varying happiness outcomes. If we failed to include those other variables, the bivariate analysis might lead to results tainted by confounding bias (which could run in either direction).

What variables should be included as controls? A safe choice is to include age and sex (Bartram 2021). There is no ambiguity about causal direction for those variables: if there are any causal relationships here, they would run from age and sex to unemployment and happiness. (Any causal effect could not go in the other direction: being unemployed cannot determine one's sex, nor how old one is, see Glenn 2009.) A similar point can be made about education: most people complete their education prior to significant job market experience. If education affects the likelihood of unemployment and also affects happiness, education is a confounder and so it must be included as a control variable for estimating the effect of unemployment on happiness. We might consider a caution in this regard: if many people who become unemployed respond by re-engaging in education, the case for including education as a control variable would be undermined.

Table 1 Exploring the impact of unemployment on happiness (OLS regression models)

	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
	b	p	b	p	b	p	b	p	b	p	b	p
Unemployed	−0.68	0.000	−0.66	0.000	−0.66	0.000	−0.37	0.000	−0.66	0.000	−0.52	0.000
Age			−0.06	0.000	−0.06	0.000	−0.06	0.000	−0.06	0.000	−0.09	0.000
Age-squared			0.00	0.000	0.00	0.000	0.00	0.000	0.00	0.000	0.00	0.000
Sex (female)			0.09	0.109								
Education (ref: primary):												
Lower secondary			−0.08	0.486	−0.07	0.543	−0.21	0.073	−0.05	0.673	−0.11	0.345
Upper secondary (lower)			0.05	0.686	0.05	0.690	−0.06	0.609	0.08	0.523	−0.00	0.995
Upper secondary (upper)			0.04	0.754	0.04	0.708	−0.13	0.235	0.07	0.536	−0.02	0.888
Advanced vocational			0.04	0.685	0.05	0.636	−0.19	0.092	0.07	0.509	0.02	0.821
Lower tertiary			0.33	0.003	0.34	0.003	0.03	0.764	0.34	0.003	0.28	0.012
Higher tertiary			0.31	0.008	0.32	0.006	−0.04	0.749	0.30	0.009	0.28	0.014
Other			0.22	0.372	0.21	0.385	0.05	0.827	0.24	0.311	0.18	0.441
Income							0.11	0.000				
Religiosity									0.04	0.000		
Partner											0.75	0.000
Constant	7.53	0.000	8.50	0.000	8.17	0.000	8.41	0.000	8.82	0.000		0.000
N	4045		4045		4045		4045		4045		4045	
R^2	0.013		0.021		0.044		0.026		0.057			

All models include dummy variables for survey round (5 through 8, with 5 as the reference category) and sample (design) weights

A model (2) that adds sex, age (and age-squared), and education gives a smaller coefficient for the effect of unemployment: −0.66, as against −0.68. There is no reason to think this model is defective in comparison to Model 1. The estimate might be biased via omission of other potentially confounding variables, but it is not biased via incorrect inclusion of sex, age, and education. However, it appears to be unnecessary to include sex: results for the unemployment variable do not change at all when it is excluded (Model 3), and so it is not a relevant confounder with respect to the impact of unemployment on happiness. Including it is not an error (particularly because it is not an intervening variable), but it is not necessary. Subsequent models therefore exclude it.

Model 4, which adds income, is another story entirely. Here, the coefficient for unemployment has been cut almost in half, down to −0.37. A model of this sort is sometimes interpreted as follows: unemployment has a direct impact on unhappiness even when we account for the indirect impact running through income. Unemployment reduces income, which reduces happiness—but there is a direct impact of unemployment as well. Statements of that sort are correct (as far as they go). However, what is incorrect is to state that "the impact" of unemployment on happiness is a reduction of 0.37 points. The impact of unemployment on happiness includes the consequent loss of income, which reduces happiness. If we control for income, we omit that indirect impact from our estimate. Considered from another angle: if we compare the unemployed to the employed while controlling income (i.e., the comparison is done across people who have the same level of income), we imply that unemployment does not change one's income – an implication that is obviously incorrect.[3]

One might expect that few researchers who investigate the relationship between unemployment and happiness are likely to construct their analyses in that way—but see e.g. Böckerman and Ilmakunnas (2006); van der Meer (2014) and Chen and Hou (2019). In any event, practices of that type abound in published (and submitted) research on happiness, exploring other topics. The practice gives substance to the idea (introduced above) that it is sometimes incorrect to say that a coefficient gives an effect that is net of the impact of "other" variables. In this instance, the coefficient of 0.37 for unemployment is not only net of the impact of the other variables in the model (age, education, and income)—it is also net of part of the impact of *unemployment itself*, i.e. the portion of unemployment's impact that runs indirectly to happiness via income.

Controlling for income, then, is a clear error (unless great/uncharacteristic care is taken with interpretation, pointing to a direct effect). What about "other determinants" of happiness? Religiosity is usually significant, statistically and substantively (e.g. Eichhorn 2012). Should it be included in a model designed to evaluate the consequences of unemployment for happiness? Consider results from Model 5 (note that this model excludes income): the coefficient for unemployment is now −0.66, identical to the value in Model 3. Religiosity is significantly associated with happiness—but it is apparently irrelevant to the question of whether unemployment affects happiness. As with sex, including religiosity is not obviously incorrect, because it is not apparent that becoming unemployed leads people to become more religious (i.e., it does not appear to be an intervening variable in this context). But by the same token it does not seem that being more religious leads to a greater (or

---

[3] In a more structural approach, researchers might consider the role played by unemployment insurance (UI). If the availability (and cost) of UI influences employers' decisions about employment and redundancy (while also affecting the happiness of those who become unemployed), then UI is a potential confounder. This sort of thinking could be important especially for research involving multiple countries.
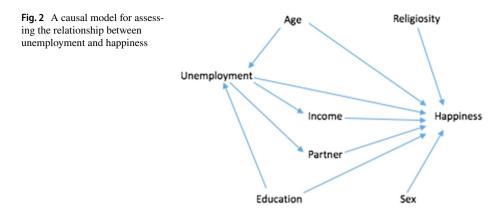
lower) likelihood of becoming unemployed (so, it is not a confounder). Inclusion of religiosity in this analysis, while not incorrect, is unnecessary (equivalent to Fig. 1.c above). As Lieberson (1985) notes, it is safest to include control variables when it emerges that it is unnecessary to do so; when including controls makes a difference, we then have further work to do, to determine *why* they make a difference. In that connection, we should draw on a theory of some sort to guide decisions on controls: should we *expect* that religiosity is a confounder, because we have reason to believe that it is a determinant of unemployment? In other words, we need to consider not only the "other determinants" of the outcome (happiness) but the determinants of the independent variable whose impact we want to identify.

What about the partner variable? This variable as well is part of many researchers' "standard set" of controls. Model 6 suggests that including it alters the estimation of unemployment's impact on happiness: the coefficient is now lower, at –0.52. Should this estimate be preferred over the model giving a coefficient of –0.66? The reduction is not trivial: it amounts to more than 21 per cent relative to the coefficient in Model 2.[4] As before, answering the question depends on our understanding (ideally, rooted in previous research) about the nature of the association between partnership and unemployment. If people living with partners are more (or less) likely to become unemployed (concisely: partner→unemployment), then the partner variable should be included as a confounder. The decision to sack someone is made by a manager; for partner→unemployment to hold, it might be the case that managers target people with (or without) partners disproportionately in determining whom to sack (or indeed whom to hire). But I am unaware of research indicating that this angle is a feature of redundancy decisions in countries like the UK in the present period.

Consider the opposite proposition: unemployment sometimes leads to relationship breakdown (concisely: unemployment→[loss of] partner). If this is our understanding of the relationship between unemployment and partner, then the partner variable should not be included: the partner variable would intervene in the path from unemployment to happiness, and controlling for it would obscure part of the impact of unemployment on happiness (just as in the income example above). The choice between these scenarios should be grounded in theory (e.g. existing research on that specific topic). Existing research offers ample support for the latter proposition (e.g. Hansen 2005; Sander 1992). I will therefore treat it as an intervening variable.

There is of course a real possibility that the situation is not as clear-cut as that: perhaps the relationship between unemployment and partnership runs in both directions. In situations of that sort, we might decide that conventional cross-sectional models are not a suitable method. We might also present readers with both results – the coefficient (for the main variable of interest) with the control included, vs. the coefficient with the control excluded. If the difference between them is large (ideally, specified in reference to a standard of some sort), readers might have an appropriate sense of imprecision in the research findings. If the difference is small, then there is little at stake in the choice. We might also consider that more complicated patterns could succeed better in describing/capturing the social reality: there might be important relationships among the controls (not just between each control and the main independent variable) that have implications for estimation of the main coefficient. Exploration of that possibility (if we have the necessary data) could be a path to better results, as research progresses on a particular topic.

---

[4] It should matter to us whether our findings are likely to be wrong by as much as 21 per cent.

Fig. 2 A causal model for assessing the relationship between unemployment and happiness

We now see as well that the impact of control variables is apparent only via comparison of results across different models. What matters (at least from an empirical angle) is what happens to the result for the main variable of interest when controls are added. The relevant information lies in the changes for that variable (not in the coefficients for the control variables). Presenting a single model that already includes controls would fail to reveal what impact the controls had in mitigating confounding bias.

Any notion of a standard set of control variables for happiness disintegrates in light of these points. The decision of which variables to include as controls will necessarily change depending on the specific research question we pose; there is no standard set. Some variables are likely safe in any situation; other variables cannot determine age and sex, and so age and sex cannot be intervening variables.[5] But in situations where inclusion of sex makes little/no difference to the other variables in the model, there is no strong case for including it; likewise with religiosity in regard to the impact of unemployment.

To make decisions regarding control variables, researchers need to abandon the idea of a standard set and ask: what are the relationships among these variables? In particular, which ones are likely to be confounders? To answer that question, we can draw on research about the main independent variable. A typical research article contains a review of existing research about the dependent variable. What is also needed is a clear sense of the research about the *independent* variable. Drawing on these literatures, we will be in a position to answer the following question: among the determinants of the dependent variable (here, happiness), are there any variables that are also determinants of the independent variable (here, unemployment)? Those are the variables that must be included as controls. Again, the needed controls will differ depending on what is being explored.

It appears from this (admittedly stylised) analysis that the impact of unemployment on happiness can be estimated effectively via a regression model that uses age and education as the only relevant confounders (Model 3). Income and (probably) partner are intervening variables. Sex and religiosity are not intervening – but by the same token they are not obviously confounders; including them does not change the picture regarding unemployment and happiness. These points are summarised in Fig. 2, which represents the assertion that only age and education (among the variables considered here) need to be controlled as confounders – because

---

[5] For sex (and also gender), that assertion might be undermined if trans-sexuality (and transgender) become more widespread as social phenomena. For the moment, of course, these possibilities are not even reflected on most major surveys.

among the other determinants of happiness only age and education also help determine unemployment (as evident in the direction of the relevant arrows).

We might wonder whether there are other potential confounders that are not used in Table 1; if so, the result given in Model 3 could be biased, via omitted variables. But in principle we should not be troubled by the fact that Model 3 is a "small" model (by virtue of containing very few control variables). As Pearl and Mackenzie note (2018, especially p. 139), lack of clarity on what controls to include has led to a tendency to include as much as one can measure—an approach he describes as "wasteful and ridden with errors." (The key error is incorrect inclusion of intervening variables.) In any event, we should not assume that a model with few controls is likely on that basis alone to be incorrectly specified.

That proposition might seem to run against the perception that the social world is complex. It is true that the social world is complex (and, to some extent any model is an abstraction, a useful simplification of reality). But it does not follow that the way to address the complexity is always to add more variables to our models. Whether a variable is needed as a control depends on its relationship to the main independent variable. If we add an intervening variable as a control, the result is exacerbation of bias. Attempts to represent complexity in a regression model must be moderated by an understanding of how control variables work.
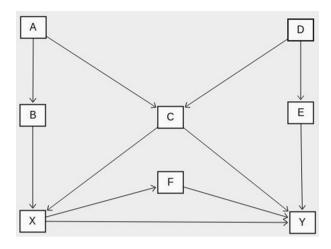
In more general terms, Fig. 2 is decidedly incomplete: it would be possible to include additional arrows, e.g. age→religiosity, or education→income. We could also add more variables, e.g. having children, which would also involve addition of arrows. It is not evident that adding complexity to the figure along these lines would lead to a different set of control variables for estimating unemployment→happiness (the examples given in this paragraph do not have that implication). Even so, development of a more general model of happiness – where relationships among the various independent variables are represented—would be a valuable research task. It should be possible to gain consensus on the control variables needed for particular research questions, and a general model could be a useful mechanism for achieving that goal. (That more ambitious task is left for future research.)

A significant caveat is required here. I have specified the causal model in Fig. 2 in part by exploring empirical results derived via analysis of data. In a more rigorous approach (e.g. Pearl and Mackenzie 2018), a model would be specified before analysis; the model itself would tell us what sort of analysis is required (i.e., what variables must—and must not—be controlled). A more rigorous approach along these lines is of course preferable. But it might pose practical dilemmas for researchers, at least at present. The assumptions we make in specifying causal models are meant to be derived from theory, i.e., existing research. But a key contention of the methodological literature forming the "causal revolution" is that a good portion of existing empirical research is potentially wrong at least to some extent, as a consequence of incorrect specification of causal models (with the consequence of mistaken decisions on control variables). In these circumstances, a more flexible and iterative approach (along the lines presented above) is advisable, at least while researchers absorb this newer perspective.

4 Towards a More Comprehensive/Systematic Approach

If researchers controlled only for confounders (and refrained from controlling for intervening variables), we would have reason to expect a substantial improvement in the quality of quantitative analysis and the findings it produces. It is however possible in principle to go further, via use of Pearl's "back-door criterion" for model specification (e.g. 1995, 2018).

Fig. 3 A prototypical causal diagram. *Note*: in this figure Y is the dependent variable, and the goal is to identify the effect of X on Y. The other variables are candidates as controls with respect to that purpose

The back-door criterion for causal identification is best appreciated via the sort of diagram (DAG) presented in Fig. 3. The usual form involves a Y (the outcome/dependent variable), an X (the main independent variable whose effect on Y we seek to identify), and other variables that might be causally related to X and/or Y. These graphs embody the assumptions we make to inform our analysis.

So, in Fig. 3 (slightly adapted from Pearl 1995), the DAG asserts that A determines B and C—but A does *not* determine D, E, F, etc. A determines X, but only through B, not directly. Moreover, B does *not* determine A; the arrow runs in only one direction. More broadly, the idea that a graph is "acyclic" means that when following the directed arrows we can pass through a variable only once. If a causal model relevant to our research question cannot be persuasively constructed in line with these criteria, we are not in a position to do causal analysis (no matter how good the data are in other respects). A key point is that omitting an arrow is a strong assumption that two variables are *not* causally related; omission requires justification no less than inclusion.

A causal effect of X on Y can be identified if we start with a correctly specified causal model and then block all the "back-door paths" from X to Y. A back-door path starts with an arrow into X and is traced across other variables until we reach Y. In Fig. 3, an example of a back-door path is: X—B—A—C—D—E—Y. Another example is: X—C—D—E—Y.

The back-door criterion works in part because back-door paths are where confounders are located. C is a confounder for the impact of X on Y, because there are arrows pointing from C to both X and Y (i.e., C is a determinant of both – as in the age/height/maths example). The arrows that run into X (forming back-door paths) help us to perceive what variables would act as confounders in this context.

Knowing how to find solutions (meeting the back-door criterion) for DAGs of this sort requires understanding an additional pattern (beyond confounders and intervening variables): colliders. A variable is a collider with respect to two other variables if arrows from those variables both point into it; in Fig. 3, C is a collider, with respect to A and D. For other patterns, we block a path by controlling for a variable on that path. Controlling for a collider, however, unblocks the path where that variable is a collider; the path is blocked as long as the collider is *not* controlled. (So, controlling for a collider typically introduces bias—see Elwert and Winship 2014.)

In Fig. 3, we must control C, to block the back-door path X—C—Y (again, C is a confounder here). But because C is a collider on the path X—B—A—C—D—E—Y, controlling it also unblocks that path. We could block that path by controlling for either B or E. A "sufficient set" for identifying the impact of X and Y, then, is either (B, C) or (C, E). There are several important implications. We do not need to control for B, C, *and* E; once C is included, either B or E is sufficient.[6] Likewise, we do not need A or D; the model suggests that they are (indirect) determinants of Y, but only via B, C, and E – so, we can identify the impact of X on Y even if it is impossible to get data on A and D. These variables might be unobserved, but this does not mean we will get results tainted by omitted variable bias. Finally: we must not control F – doing this would mean controlling for an intervening variable and would introduce bias.

These ideas merit attention from empirical researchers in general. It seems likely, however, that most scholars (outside of epidemiology, at any rate) will not find themselves immediately ready to integrate them into their research practice. Finding solutions meeting the back-door criterion (e.g. when the DAGs are even more complex) is not an obstacle; there are algorithms that can assist (e.g. Textor et al. 2011; Malinsky and Danks 2018—of course, sometimes the result is that a solution cannot in fact be identified with the available data). Instead, the difficulty will likely consist in specifying a causal model (via use of DAGs or otherwise) that researchers could defend in the way Pearl and others have in mind (on the basis of scientific knowledge). If a good many studies have been carried out incorrectly, what basis can we use to construct a model that satisfies that standard (cf. Elwert 2013)? So, I describe the more general framework only briefly in this section, for awareness; in practical terms I anticipate that the simpler propositions articulated and demonstrated above will prove more useful, at least for now.

5 Conclusion

When we specify a convincing model as described above (and in particular when others e.g. peer reviewers and readers are in fact convinced), we should be willing to interpret our results in causal terms. With grounding of this sort, our results are no longer mere correlations or associations. The well-worn aphorism "correlation is not causation" is best understood as incomplete: correlation is not causation – but it *might* be, given a proper understanding of causal models and of the specific assumptions required to interpret results effectively.[7] More advanced techniques are of course available for more robust evaluation of these relationships. What is offered here is a better way of doing cross-sectional analysis – a middle ground where the findings have some value in evaluating causal relationships.

The core principles and recommendations on offer here are usefully summarised as follows:

1. Control variables should be selected not only by virtue of being other determinants of the outcome but also on the basis of how they relate to the core independent variable

---

[6] One might wonder why the set (B, C) does not leave the path X—C—D—E—Y unblocked, if controlling for C means unblocking it as a collider. But C is a collider only with respect to the path(s) containing arrows that point into it (here: X—B—A—C—D—E—Y). So, on the path X—C—D—E—Y C is not a collider and controlling for it means blocking it (hence E is unnecessary).

[7] In Tufte's formulation (2004): "Correlation is not causation – but it sure is a hint".

whose effect we seek to gauge. An appropriate control is a *confounder* – a variable that determines not only the outcome (e.g. SWB) but also the main independent variable. Some other determinants might be irrelevant to a correct specification of a causal effect (e.g. if they do not also determine the key independent variable).

2. Variables that intervene in a path from the independent variable to the outcome must be omitted from one's model; controlling them would introduce bias. They should be introduced at a later stage only to investigate mechanisms for a causal effect identified at an earlier stage.

3. Potential controls that might operate both as confounders and intervening variables (so, the relevant arrow could run in both directions) require particular attention. Options for handling them include deciding that the relationship is more consequential in one direction than in the other – a point to be considered as a limitation in connection with presenting one's results. Or, we could simply acknowledge that identifying a causal effect in these circumstances is not feasible, at least not via cross-sectional analysis.

4. If a variable has been selected for inclusion on the basis that it makes sense as a control (i.e., because it is a confounder with respect to the main independent variable), then its coefficient cannot be interpreted in causal terms as a total effect (Westreich and Greenland 2013; Hernán 2018). If variables have been properly selected as controls with respect to a specific independent variable, that variable will always intervene in a path from the controls to the outcome. The necessary consequence is that the coefficients for the controls do not give a total effect.

5. DAGs are a useful representation of these principles, effective in conveying the information needed for the specific decisions involved in constructing one's analytical model. They work when we have sufficient clarity about one's modelling assumptions – but they can also function as a signal that we have not yet achieved sufficient clarity to conduct our empirical analysis.

If specified correctly (with clear articulation of assumptions the researcher can justify persuasively), regression models using observational data—even cross-sectional data—can then help us produce findings that have some value in evaluating possible causal effects. It is then no longer necessary to disclaim that "it's only an association". The models are however only as plausible as the assumptions—and plausibility depends on being able to make a clear distinction between confounders and intervening variables. The complexity of the social world might sometimes impede clarity along these lines (as in our discussion of the partner variable above, and summary point #3). We will also sometimes need to acknowledge that we lack the data that correspond to these assumptions.

Models specified via those criteria are likely to be quite different from models constructed according to other principles – e.g. "explaining variation" and/or prediction (Shmueli 2010). The choice of relevant criteria follows from clarity on one's research purpose—and I emphasize that what many researchers want (sometimes against long-standing misdirection from earlier training in statistics) is to evaluate causal effects relevant to their core research topics. For this purpose it is of course possible to go beyond single-stage regression models of cross-sectional data; researchers should aspire to improve their skills and use the more advanced techniques that meet a higher standard. But while on the path towards that destination, it is possible to do better cross-sectional analysis that avoids some obvious errors undermining the value of one's work.

Compliance with Ethical Standards

Appendix A note on assumptions

This article has used the word "assumptions" in ways that depart from how it is used in more conventional practice. In some discussions (statistics textbooks, in particular), a typical set of assumptions includes items such as: homoskedasticity, no overly-influential observations, no multicollinearity, a linear relationship between dependent and independent variables, normality of residuals, and independence between the error term and the independent variables (e.g. Agresti and Healey 1997). Some of these pertain only to linear regression for continuous dependent variables (though equivalent assumptions would be necessary for other functional forms, e.g. logistic regression). Others (e.g. homoskedasticity) are relevant for hypothesis tests, which I have not emphasised here.

The word assumptions as used in this article invokes a different question, arguably a more fundamental one: what do we already know (i.e., what can we assume) about what leads to certain patterns/outcomes in regard to the topic we are investigating? What do we need to take into account as we try to learn something new (e.g. about a factor we haven't previously considered)? This is the information encapsulated in a DAG. Issues like homoskedasticity are secondary, in comparison. The item from the conventional list worth considering further is "independence between the error term and the independent variables". This formulation is an obscure way of expressing the important idea that our results depend on not omitting confounding variables from our model; if we omit confounders (leaving their influence buried in the error term), we get biased results.

This idea is captured better by the notion that our results depend on a correctly specified causal model—in other words, a set of assumptions (in the meaning of that word as used here). If we construct a model that incorrectly assumes there is no relationship between a particular factor and other variables in the model (including the dependent variable), we omit a confounder from our empirical analysis. This can happen via not including a variable in a DAG; as noted, omission of arrows represents a strong assumption that there is no causal relationship between variables, and we cannot draw an arrow if the variable in

question is not in the DAG. This way of expressing the point is much more digestible than the notion that we must ensure "independence between the error term and the independent variables".

References

Agresti, A., & Finlay, B. (1997). *Statistical methods for the social sciences*. Upper Saddle River, N.J.: Prentice Hall.

Arnold, K. F., Davies, V., de Kamps, M., et al. (2020). Reflections on modern methods: generalized linear models for prognosis and intervention. *International Journal of Epidemiology*. https://doi.org/10.1093/ije/dyaa049.

Bartram, D. (2011). Identity, migration, and happiness. *Sociologie Românească, 9*(1), 7–13.

Bartram, D. (2021). Age and life satisfaction: Getting control variables under control. *Sociology*. https://doi.org/10.1177/0038038520926871.

Bell, A., Fairbrother, M., & Jones, K. (2019). Fixed and random effects models: Making an informed choice. *Quality and Quantity, 53,* 1051–1074. https://doi.org/10.1007/s11135-018-0802-x.

Berk, R. A. (1983). *Regression analysis: A constructive critique*. Thousand Oaks: Sage Publications.

Berk, R. A. (2010). What you can and can't properly do with regression. *Journal of Quantitative Criminology, 26*(4), 481–487. https://doi.org/10.1007/s10940-010-9116-4.

Böckerman, P., & Ilmakunnas, P. (2006). Elusive effects of unemployment on happiness. *Social Indicators Research, 79*(1), 159–169. https://doi.org/10.1007/s11205-005-4609-5.

Chen, W.-H., & Hou, F. (2019). The effect of unemployment on life satisfaction: A cross-national comparison. *Applied Research in Quality of Life, 14,* 1035–1058. https://doi.org/10.1007/s11482-018-9638-8.

Dieleman, J. L., & Templin, T. (2014). Random-effects, fixed-effects and the within-between specification for clustered data in observational health studies: A simulation study. *PLoS ONE, 9*(10), e110257. https://doi.org/10.1371/journal.pone.0110257.

Eichhorn, J. (2012). Happiness for believers? Contextualizing the effects of religiosity on life-satisfaction. *European Sociological Review, 28*(5), 583–593. https://doi.org/10.1093/esr/jcr027.

Elwert, F. (2013). Graphical causal models. In S. L. Morgan (Ed.), *Handbook of causal analysis for social research* (pp. 245–273). Dordrecht: Springer.

Elwert, F., & Winship, C. (2014). Endogenous selection bias: The problem of conditioning on a collider variable. *Annual Review of Sociology, 40*(1), 31–53. https://doi.org/10.1146/annurev-soc-071913-043455.

Ferrer-i-Carbonell, A., & Frijters, P. (2004). How important is methodology for the estimates of the determinants of happiness? *The Economic Journal, 114*(497), 641–659. https://doi.org/10.1111/j.1468-0297.2004.00235.x.

Freedman, D. A. (1991). Statistical models and shoe leather. *Sociological Methodology, 21,* 291–313.

Gangl, M. (2010). Causal inference in sociological research. *Annual Review of Sociology, 36*(1), 21–47. https://doi.org/10.1146/annurev.soc.012809.102702.

Glenn, N. (2009). Is the apparent U-shape of well-being over the life course a result of inappropriate use of control variables? A commentary on Blanchflower and Oswald. *Social Science and Medicine, 69*(4), 481–485. https://doi.org/10.1016/j.socscimed.2009.05.038.

Glynn, A. N., & Kashin, K. (2017). Front-door difference-in-difference estimators. *American Journal of Political Science, 61*(4), 989–1002. https://doi.org/10.1111/ajps.12311.

Gonza, G., & Burger, A. (2017). Subjective well-being during the 2008 economic crisis: Identification of mediating and moderating factors. *Journal of Happiness Studies, 18*(6), 1763–1797. https://doi.org/10.1007/s10902-016-9797-y.

Hansen, H.-T. (2005). Unemployment and Marital dissolution. *European Sociological Review, 21*(2), 135–148. https://doi.org/10.1093/esr/jci009.

Hernán, M. A. (2018). The C-word: scientific euphemisms do not improve causal inference from observational data. *American Journal of Public Health, 108*(5), 616–619. https://doi.org/10.2105/AJPH.2018.304337.

Hou, F. (2014). Keep up with the joneses or keep on as their neighbours: Life satisfaction and income in canadian urban neighbourhoods. *Journal of Happiness Studies, 15*(5), 1085–1107. https://doi.org/10.1007/s10902-013-9465-4.

Jowell, R. (2007). *European social survey, technical report*. London: Centre for Comparative Social Surveys, City University.

Lieberson, S. (1985). *Making it count: The improvement of social research and theory*. Berkeley: University of California Press.

Malinsky, D., & Danks, D. (2018). Causal discovery algorithms: A practical guide. *Philosophy Compass, 13*(1), e12470. https://doi.org/10.1146/annurev.psych.58.110405.085542.

Morgan, S. L., & Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles for social research*. Cambridge: Cambridge University Press.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika, 82*(4), 669–688. https://doi.org/10.1093/biomet/82.4.669.

Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics Surveys, 3,* 96–146. https://doi.org/10.1214/09-SS057.

Pearl, J. (2010). An introduction to causal inference. *The International Journal of Biostatistics*. https://doi.org/10.2202/1557-4679.1203.

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. London: Allen Lane.

Rohrer, J. M. (2018). Thinking clearly about correlations and causation: graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science, 1*(1), 27–42. https://doi.org/10.1177/2515245917745629.

Sander, W. (1992). Unemployment and marital status in Great Britain. *Biodemography and Social Biology, 39*(3–4), 299–305. https://doi.org/10.1080/19485565.1992.9988825.

Shahar, E., & Shahar, D. J. (2013). Causal diagrams and the cross-sectional study. *Clinical Epidemiology, 5,* 57–65. https://doi.org/10.2147/CLEP.S42843.

Shmueli, G. (2010). To explain or to predict? *Statistical Science, 25*(3), 289–310. https://doi.org/10.1214/10-STS330.

Smith, H. L. (1990). Specification problems in experimental and non-experimental social research. *Sociological Methodology, 20,* 59–91.

Textor, J., Hardt, J., & Knüppel, S. (2011). DAGitty: A graphical tool for analyzing causal diagrams. *Epidemiology, 22*(5), 745.

Tufte, E. R. (2004). *The cognitive style of powerpoint*. Cheshire, Conn: Graphics Press.

van der Meer, P. H. (2014). Gender, unemployment and subjective well-being: why being unemployed is worse for men than for women. *Social Indicators Research, 115*(1), 23–44. https://doi.org/10.1007/s11205-012-0207-5.

VanderWeele, T. J., & Shpitser, I. (2011). A new criterion for confounder selection. *Biometrics, 67*(4), 1406–1413. https://doi.org/10.1111/j.1541-0420.2011.01619.x.

Westreich, D., & Greenland, S. (2013). The table 2 fallacy: Presenting and interpreting confounder and modifier coefficients. *American Journal of Epidemiology, 177*(4), 292–298. https://doi.org/10.1093/aje/kws412.

Wooldridge, J. M. (2005). Violating ignorability of treatment by controlling for too many factors. *Econometric Theory, 21*(5), 1026–1028. https://doi.org/10.1017/S0266466605050516.

York, R. (2018). Control variables and causal inference: a question of balance. *International Journal of Social Research Methodology, 21*(6), 675–684. https://doi.org/10.1080/13645579.2018.1468730.