Published in final edited form as:

Perspect Psychol Sci. 2010 October 11; 5(5): 546–556. doi:10.1177/1745691610383511.

Causal Inference and Observational Research: The Utility of Twins

Matt McGue^{1,2}, Merete Osler^{2,3}, and Kaare Christensen⁴

¹ Department of Psychology, University of Minnesota, Minneapolis

² Institute of Public Health, University of Southern Denmark, Odense, Denmark

³ Research Center for Prevention and Health, Glostrup Hospital, Glostrup, Denmark

⁴ The Danish Twin Registry and The Danish Aging Research Center Institute of Public Health, University of Southern Denmark, Odense, Denmark

Abstract

Valid causal inference is central to progress in theoretical and applied psychology. Although the randomized experiment is widely considered the gold standard for determining whether a given exposure increases the likelihood of some specified outcome, experiments are not always feasible and in some cases can result in biased estimates of causal effects. Alternatively, standard observational approaches are limited by the possibility of confounding, reverse causation, and the nonrandom distribution of exposure (i.e., selection). We describe the counterfactual model of causation and apply it to the challenges of causal inference in observational research, with a particular focus on aging. We argue that the study of twin pairs discordant on exposure, and in particular discordant monozygotic twins, provides a useful analog to the idealized counterfactual design. A review of discordant-twin studies in aging reveals that they are consistent with, but do not unambiguously establish, a causal effect of lifestyle factors on important late-life outcomes. Nonetheless, the existing studies are few in number and have clear limitations that have not always been considered in interpreting their results. It is concluded that twin researchers could make greater use of the discordant-twin design as one approach to strengthen causal inferences in observational research.

Keywords

discordant-twin design; causal inference; twin research; lifestyle influences in aging

A central aim within psychological gerontology is to explicate the causal mechanisms that underlie the aging of psychological functions. Without such understanding, the development of effective interventions is likely to be impeded; our understanding of psychological aging processes will be incomplete (Depp & Jeste, 2006; Rowe & Kahn, 1987, 1999). Thus, it is important that aging researchers determine whether social engagement protects against depression (Cacioppo, Hughes, Waite, Hawkey, & Thisted, 2006; H. Christensen et al., 1996; Glass, De Leon, Bassuk, & Berkman, 2006; Salthouse, 2006), whether intellectual

© The Author(s) 2010

Corresponding Author: Matt McGue, Department of Psychology/Elliott Hall, University of Minnesota, Minneapolis, MN 55455, mcgue001@umn.edu.

Reprints and permission: sagepub.com/journalsPermissions.nav

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

stimulation reduces risk of cognitive impairment (H. Christensen et al., 1996; Salthouse, 2006), and whether a physically active lifestyle increases longevity while reducing morbidity (Trost, Owen, Bauman, Sallis, & Brown, 2002). Nonetheless, as in many areas in psychology, the standard methodological approaches available to aging researchers severely constrain their ability to draw unambiguous causal inferences. The purpose of this article is to both characterize some of the major challenges associated with causal inference in observational psychology, with a particular focus on psychological gerontology, and describe how a twin study, long a staple in behavioral genetics, might address some of these challenges. Our article begins with a discussion of the nature of causation and alternative approaches to estimating causal effects. We next discuss how twin studies might be used to strengthen causal inference in observational research, and review how twin studies have been used in this way in psychological gerontology. Finally, we report an application of a twin-study design to a longstanding controversy: Does moderate alcohol consumption promote, interfere with, or have no effect on late-life cognitive functioning?

What is a Causal Effect and How Can It Be Estimated?

The Counterfactual Model of Causality

One of the enduring debates within the philosophy of science concerns the nature of cause (Parascandola & Weed, 2001; Pearl, 2000). Rather than describe alternative philosophical approaches to defining what is meant by cause, we can highlight the essence of the issue by considering what is meant when someone claims that social disengagement is a cause of dementia. First, they are probably not claiming that no socially engaged person ever develops dementia. Similarly, they are unlikely to be arguing that every disengaged person eventually becomes demented. Rather, the conception of causation within the social and behavioral sciences extends beyond formal deterministic relationships based on sufficiency and necessity. In this regard, the statement probably comes closest to meaning something like, if an individual becomes socially engaged he or she will lower their probability of developing dementia below what it would have been had they not been engaged. This perspective on causation allows that, even though on average individuals would reduce their chance of becoming demented if they increased their level of social engagement, many socially engaged individuals will develop dementia, whereas many socially disengaged individuals will not. But how can this conception of causation be formalized and how might we use this definition to evaluate alternative methodologies for drawing causal inferences within the behavioral and social sciences?

Recently, Rubin and colleagues (Rubin, 2001, 2007, 2008), building on the earlier work by Neyman (Neyman, Dabrowska, & Speed, 1923/1990), proposed a framework for evaluating causal inference within epidemiology. The counterfactual or potential outcomes framework provides both a coherent definition of causal effect for complex multidetermined outcomes (Hernán, 2004) and an integrative framework for evaluating alternative approaches for the estimation of causal effects (Little & Rubin, 2000). Suppose we are interested in the effect of some exposure X ($= T$ or C) on outcome Y in a population of N individuals. We would then define the causal effect of exposure X for the i th person as $\delta_i = Y_i^T - Y_i^C$, where Y_i^T is the i th person's outcome when exposed and Y_i^C is the outcome when he or she is not exposed. The average causal effect is obtained by taking the average of the individual effects, or $\bar{\delta} = \bar{Y}^T - \bar{Y}^C$. Of course it is not possible to observe an individual's outcome both when that individual is exposed and at the same time when he or she is not. Consequently, causal effects so defined cannot be directly observed. Rather, the virtue of the counterfactual model is that it places the problem of causal inference within a general missing data framework. That is, valid causal inference requires approaches that allow us to accurately estimate the missing nonexposure outcomes for those who were exposed and, conversely, the missing

exposure outcomes for those who were not exposed. It is important to note that there is no fundamental difference between experimental and observational approaches—both are evaluated in terms of whether they allow for accurate estimation of the relevant counterfactuals.

There is general recognition that experimental approaches, being based on the ability to randomly assign individuals to different levels of exposure, afford the greatest opportunity for strong causal inference (West et al., 2008). The counterfactual model helps us to understand why. Specifically, random assignment should ensure that those assigned to the control group will provide on average a reasonable counterfactual for those exposed to the treatment and, conversely, that those assigned to the treatment should, on average, provide a reasonable counterfactual for those who were not. The counterfactual approach also helps to identify some of the limitations of experimental approaches for causal inference. For

example, even with random assignment, the estimated average causal effect, $\bar{\delta}$, may be substantially biased if treatment response among those who volunteer to participate in an experiment differs from those who decline to participate. It seems almost certain that those who participate in interventions that provide, for example, high levels of intellectual, social, and physical stimulation will not be representative of the general population of older individuals (Elzen, Slaets, Snijders, & Steverink, 2008; van Heuvelen et al., 2005). Loss to follow-up and failure to comply with intervention guidelines are additional factors that can lead to biased estimates of causal effects in treatment studies. Although clearly powerful, experimental approaches may systematically underestimate or overestimate the effect of social and behavioral interventions (Levitt & List, 2007; Ross, Thomsen, Boesen, & Johansen, 2004). Inferences from experimental studies should be validated with alternative approaches. Moreover, for many important psychological questions, an experiment may not be feasible or ethical or it may not adequately address the central question of interest. For example, the effects of a short-term intensive physical or intellectual intervention may tell us little about the impact of lifestyle factors whose effects play out over years rather than weeks (Rejeski et al., 2005). Despite their perceived strengths, experiments may not always be the best approach to causal inference in the behavioral sciences.

Causal Inference and Observational Research

Concern with observational approaches to causal inference center on two alternatives to true causation as an explanation for an association between exposure and outcome: reverse causation and confounding. Both can be illustrated by considering the association that exists between intellectual engagement and cognitive functioning and the controversy that surrounds its interpretation. This association may arise because of true causation (i.e., intellectual engagement influences cognitive functioning) as some have suggested (Mackinnon, Christensen, Hofer, Korten, & Jorm, 2003; Pushkar et al., 1999). Alternatively, the association may also arise by reverse causation, whereby high cognitive functioning leads to intellectual engagement as others have suggested (Hertzog, Hultsch, & Dixon, 1999; Salthouse, 2006). In principle, resolving true causation from reverse causation can be achieved through longitudinal research that allows for the sequencing of exposure and outcome. In practice, the valid sequencing of exposure and outcome is far from trivial, involving significant design challenges and strong assumptions about the nature of measurement (Rogosa, 1980).

Arguably, confounding represents a more pervasive challenge to valid inference in observational research than the possibility of reverse causation. For a variable to be considered a confounder it must (a) be a cause of the outcome, (b) be correlated with exposure, and (c) not be affected by exposure (McNamee, 2003). In the present case, educational attainment might be considered an example of a potential confounder, because

educational attainment might be a cause of cognitive functioning (the outcome) and correlated with, but not caused by, an intellectually engaged lifestyle (the exposure). Researchers have generally attempted to deal with confounding in observational research either through post hoc statistical adjustments (e.g., covariate adjustment) or by executing research designs that approximate a randomized trial even if observational (e.g., the use of natural experiments; Rubin, 2008).

Unfortunately, post hoc statistical approaches to confounding are fraught with difficulty (Weinberg, 1993). Most psychologists are familiar with the use of partial correlational methods to correct for potential confounders. They will also be familiar with the limitations of this approach (or its regression-based equivalents). Adjustment cannot be made for confounders that were not assessed, and adjustment for confounders that were measured poorly is certain to be inadequate. Moreover, only rarely do social and behavioral scientists rigorously evaluate whether the covariates in their regression model meet the criteria for a confounder (Meehl, 1971). As a consequence, even if it seems sensible, adjusting the correlation between cognitive functioning and intellectual engagement for educational attainment will not provide an unbiased estimate of causal effect if educational attainment is not a true confounder (e.g., because it is consequence rather than a cause of cognitive functioning), if educational attainment is measured with error (in which case the statistical adjustment would be imperfect), or if a key confounder was not included as a covariate (e.g., the association of intellectual engagement and cognitive functioning may owe in part to common genetic factors).

Some social and behavioral scientists have turned to structural equation modeling (SEM) approaches in an attempt to address the problems of confounding (Anderson & Gerbing, 1988; Bentler, 1985). There is no doubt that SEM has gained enormously in popularity over the past few decades. The number of entries to SEM in social and behavioral sciences journals in Thomson's Institute for Scientific Information database increased from 2 in 1977 to over 900 in 2008. Yet although there is little question over the utility of SEM for describing complex multivariate data sets such as those generated in longitudinal research (Tomarken & Waller, 2005), less certain is the extent to which SEM has facilitated valid causal inference by addressing the problem of confounding in observational data (Freedman, 2006).

A second general approach for dealing with confounding in observational research is the development of research designs that approximate those of a true experiment (Shadish, Cook, & Campbell, 2002). Rutter (2007) recently provided a broad overview of natural experiments available to observational researchers in the social and behavioral sciences. Although these approaches are not without their own limitations, the advantage of natural experiments is that they approach the problem of confounding as something to be addressed a priori through research design rather than a posteriori through statistical analysis. Our focus here is on one of the natural experiments described by Rutter (2007). The cotwin control or discordant-twin-pairs design has the advantage that there are a large number of twin studies in psychology even if the particular use of a twin study we will emphasize here is not often taken advantage of.

The Cotwin Control or Discordant-Twin-Pairs Design

Logic of the Cotwin Control Design

Twin studies have long been a staple within the field of behavioral genetics. There are hundreds, perhaps thousands of twin studies of behavioral and social phenotypes (Bouchard & Propping, 1993; McGue & Bouchard, 1998). In most cases, these studies have sought to estimate heritability by comparing the phenotypic similarity of monozygotic (MZ) twins,

who are genetically identical, with that of dizygotic (DZ) twins, who are no more genetically similar than ordinary siblings. Less widely recognized perhaps is the utility of twin studies to explore the environmental basis of individual differences in behavior (Kaprio et al., 1993). Specifically, twins discordant for exposure approximate the alternative outcomes presumed under the counterfactual model. The basic logic of the approach can be illustrated by one of the earliest applications of the so-called cotwin control or discordant-twin design: Does smoking increase mortality?

The eminent statistician, R.A. Fisher, was skeptical that the association of smoking with lung cancer reflected a causal effect of the former on the latter: “that cigarette-smoking and lung cancer, though not mutually causative, are both influenced by a common cause, in this case the individual genotype” (Fisher, 1957, p. 298). For Fisher, genetic confounding rather than true causation might be the basis of the association of smoking with lung cancer. In a subsequent letter to *Nature*, Fisher (1958) provided what he thought to be support for his genetic hypothesis: “that the smoking habits of monozygotic, or one-egg, twins were clearly more alike than those of twins derived from two eggs” (p. 596). But heritability of exposure does not preclude causality of effect. Fisher did not undertake the critical follow-up to his observation that smoking exposure was heritable. Namely, among twins who are discordant for smoking, is the smoking member of the pair more likely to suffer lung cancer and early mortality than the nonsmoking member? Subsequent research on MZ and DZ twin pairs discordant for smoking has adequately answered this question. Within discordant pairs, the smoking twin has significantly increased mortality relative to the nonsmoking twin (Carmelli & Page, 1997; Kaprio & Koskenvuo, 1989). Findings from the discordant-twin design led to the disconfirmation of Fisher’s genetic hypothesis by providing more powerful support for a causal influence of smoking than is possible with standard epidemiological designs.

It is helpful to analyze the logic of the cotwin control approach within the counterfactual framework. Let Y_i^S be the outcome when the i th twin is a smoker and Y_i^{NS} be the outcome when he or she is not. Then the causal effect of smoking for the i th twin is given by $\delta_i = Y_i^S - Y_i^{NS}$. Assume further that $Y_{i'}^{NS}$ is the outcome for the i th twin’s nonsmoking cotwin (designated i'). Then the cotwin control design involves estimating δ_i by $\widehat{\delta}_i = Y_i^S - Y_{i'}^{NS}$. That is, the cotwin control design involves implicitly using the nonsmoking cotwin of a smoker to estimate what the smoker would have looked like had he or she not smoked. Of course the nonsmoking cotwin will not necessarily be a perfect match under this alternative outcomes scenario. Nonetheless, discordant MZ twins, who have both a common genotype and a common early rearing environment, are likely to be better matches than discordant DZ twins, who have a common early rearing environment but share only 50% of their segregating genetic alleles.

These expectations can be formalized within a biometric formulation. That is, the observed phenotype (P) or outcome can be assumed to be a function of (a) additive genetic effects (A), which are shared completely by MZ twins but are only 50% shared by DZ twins; (b) shared early environmental effects (C ; e.g., the effects of rearing social class), which are shared completely by both MZ and DZ twins; and (c) non-shared environmental effects (E ; e.g., the effects of adult social class), which are not shared by either MZ or DZ twins. Individual-level associations (e.g., as estimated from an individual-level regression of outcome on exposure) reflect potential confounding of exposure and outcome by A , C , and E effects. Associations within DZ twin pairs discordant for exposure (i.e., the exposure effect estimated comparing the two members of discordant DZ pairs), control for C effects and partially for A effects. Associations within MZ twin pairs discordant for exposure control for both C and A effects. Neither within-pair comparison controls for confounding due to E

factors, an important point to which we return below. Regardless, if exposure is truly a cause of outcome, then we expect exposure to be associated with outcome at the individual level, as well as within DZ twin pairs discordant for exposure and within MZ twin pairs discordant for exposure.

Figure 1 illustrates three possible patterns of individual-level versus within-pair effects (cf. Bergen et al., 2008). In Scenario A, the effect of exposure is the same at the individual level and within DZ and MZ pairs discordant on exposure. This is the pattern we would expect if exposure were causal and there was no confounding by either A or C. In Scenario B, the within-pair effect is reduced in DZ twins relative to the individual-level effect and is completely absent in MZ twins. This pattern suggests that the association of exposure and outcome is due entirely to confounding by A and C. In this case, failure to observe an effect within discordant MZ twin pairs is inconsistent with causality. In Scenario C, the exposure effect is reduced but not eliminated within both DZ and MZ pairs discordant on exposure relative to the individual-level effect. This pattern suggests that the association of exposure to outcome is due partially to confounding, whereas the existence of an exposure effect within discordant MZ pairs is consistent with at least a partial causal effect. Clearly the power of the discordant-twin design stems largely from the within-MZ pair analysis, because MZ comparisons afford the greatest control for confounding. Nonetheless, the biometric formulation also helps to identify one of the limitations of the discordant-twin approach. That is, only exposures on which MZ twins differ can be explored (i.e., exposures mediated by the non-shared environment), so that this design cannot be of help in exploring the impact of shared environmental exposures such as maternal depression and rearing social class—at least in the usual case when the MZ twins have been reared together.

Discordant-Twin Studies in Gerontology

The extent to which the discordant-twin design will be of utility in gerontology will depend on the degree to which exposure to putative aging risk factors is heritable, just as Fisher (1958) reasoned for smoking more than 50 years ago. That is, the power of the discordant-twin design is that it controls for potential genetic (and also shared environmental) confounding, and without heritability there can be no genetic confounding. In fact, twin studies have shown that a wide range of factors of relevance to aging are heritable. These include smoking (Li, Cheng, Ma, & Swan, 2003), heavy alcohol consumption (Whitfield et al., 2004), achieved social class (Plomin & Bergeman, 1991), social engagement (McGue & Christensen, 2007), and physical activity (Frederiksen & Christensen, 2003). The heritability of these lifestyle factors may reflect selection effects, whereby individuals adopt lifestyles that complement and reinforce heritable characteristics (Scarr & McCartney, 1983). We know, for example, that genetic factors contribute to individual differences in personality (Finkel & McGue, 1997). We know further that individuals with outgoing and cheerful dispositions are more likely to have satisfying relations and a socially engaged lifestyle than are individuals who are withdrawn or sullen, in part because cheerful individuals strive to construct experiences that reinforce their sunny dispositions (Ozer & Benet-Martinez, 2006). Thus, heritable influences on personality become translated as heritable influences on social engagement, raising the possibility that the association of social engagement with, for example, depression risk reflects genetic selection rather than true causation. But as with smoking exposure, the mere demonstration of heritability is not sufficient to establish genetic selection. If the association of social engagement with outcome is due entirely to genetic selection, then differences within genetically identical MZ twin pairs discordant for engagement should be unrelated to differences in outcome.

Table 1 summarizes findings from discordant-twin studies in the psychological aging field in terms of the association between exposure and outcome at the individual level (i.e., without regard to twin-pair status) and the various within-pair estimates reported in each

study. Although the number of studies is limited, a range of exposures and outcomes has been investigated. Several tentative conclusions seem appropriate. First, for most outcomes, findings from individual-level analyses are confirmed by showing exposure effects within discordant twin pairs, even if the magnitudes of the within-pair effects are smaller than those at the individual level. The one notable exception is adult social class, for which Osler, McGue, and Christensen (2007) reported no association with health and cognitive outcomes in MZ twin pairs discordant for adult social class. Second, in most cases, the discordant-twin-pair analysis did not distinguish MZ from DZ twins, making it difficult to interpret the significance of the results (cf. Fig. 1). Failure to report findings by zygosity presumably owes to the modest size of most of the discordant twin samples. Nonetheless, a pooled MZ and DZ analysis cannot totally rule out the possibility of genetic confounding since an association within discordant DZ pairs, who are not matched for genotype, is expected even under genetic selection. Finally, in the few studies that analyzed discordant MZ twins, findings generally confirm those found at the individual level. Thus, Carlson, Andersson, Lichtenstein, Michaelsson, and Ahlbom (2007) reported that among MZ twins discordant for physical activity, the less active twin had higher mortality over a 29-year period; Kujala, Kaprio, and Koskenvuo (2002) reported that smoking and heavy alcohol use were associated with mortality within MZ twin pairs discordant on exposure; and McGue and Christensen (2007) reported that in MZ twin pairs discordant for social engagement, the least social twin had lower physical and cognitive functioning and higher rates of depression symptomatology.

It is important to recognize that the comparison of discordant MZ twins does not guarantee certain causal inference even if it is more powerful than individual-level analyses. The generalizability of findings from research on twins will always be questioned, even though research has repeatedly shown that twins are generally unremarkable with respect to their personalities (Johnson, Krueger, Bouchard, & McGue, 2002), cognitive abilities (K. Christensen et al., 2006), risk for mental disorders (Kendler, Pedersen, Farahmand, & Persson, 1996), and adult mortality trajectories (K. Christensen, Vaupel, Holm, & Yashin, 1995). In addition, the standard discordant-twin design cannot rule out reverse causation even if it addresses issues surrounding confounding. That is, if outcome differences lead to differences in exposure (e.g., as might occur when the outcome is cognitive ability and exposure is intellectual engagement), then we would still expect to observe within-pair associations. Ruling out reverse causation requires longitudinal designs.

A further significant limitation concerns the asymmetrical nature of causal evidence from a discordant-twin study. That is, although failure to observe a within-pair association would seem to disconfirm the existence of a causal effect (although even this can be debated; see discussion on reliability below), the observation of an association between exposure and outcome within discordant pairs cannot establish causality. A recent example from the substance abuse field can help to illustrate the issue. Lynskey et al. (2003) reported that among MZ twins discordant for early onset cannabis abuse, the early-using twin was more likely to be a drug abuser or an alcoholic as an adult than the late-using or nonusing twin. These findings are consistent with, but do not establish, a causal effect of early cannabis use on adult substance abuse. This is because the factors that led to the twins being discordant in their cannabis use might also be the factors that account for their differences in outcome. This may be the case with early cannabis use. In a separate study, Vink, Nawijn, Boomsma, and Willemsen (2007) reported that MZ twin discordance in cannabis exposure was associated with preexisting personality indicators of risk. That is, the early-using twin was more likely to be high in sensation seeking and neuroticism than his or her nonearly-using but genetically identical cotwin. Consequently, the association of early cannabis use with adult outcomes, rather than being causal, may arise because the personality factors that increase risk of early use also increase risk of adult substance abuse.

Thus the discordant-twin design does not obviate the need to consider the role of confounding factors even though it provides a basis for powerful tests of causality in observational settings. This is because MZ twins do not provide a perfect counterfactual pair. That is, even if they are matched on genotype and early rearing environment, the two members of an MZ twin pair are not matched on the nonshared experiences that make them psychologically unique. Consequently, a within-MZ pairs association of exposure with outcome may reflect true causality or, alternatively, the effect of the nonshared experiences that led to differences in exposure. Unfortunately, none of the studies included in Table 1 systematically explored the factors associated with twin discordance in exposure.

Finally, in some cases the discordant-twin design may fail to observe a within-pair association even when there is a causal influence of exposure on outcome and the sample is large enough to provide adequate power. Specifically, measurement error in the exposure variable is expected to attenuate within-pair associations to a greater degree than individual-level associations because of the compounding of error inherent in taking a difference score (Ashenfelter & Krueger, 1994). Under classical test score theory, if σ is the common estimate of the proportion of measurement error in the exposure variable and ρ is the twin correlation, then the regression coefficient of outcome on exposure is attenuated by σ at the individual level but by $\sigma/(1-\rho)$ at the within-pair level (Griliches, 1979; note that we implicitly assume here that there is no error of measurement in the outcome, because if it exists it would not have a differential effect on the alternative estimates discussed here). Thus, for example, if measurement error accounted for 10% in an exposure variable, then the individual-level estimate would be attenuated by 10%, but the within-pair estimate would be attenuated by 33% when the twin correlation on exposure is .7 and 20% when the twin correlation is .5. Consequently, we expect attenuation of within-pair estimates due to the compounding of measurement error. Moreover, since MZ twin correlations are typically higher than DZ twin correlations, we expect the within-pair attenuation to be greater for MZ than DZ pairs.

An Example of the Discordant-Twin Design: Moderate Drinking and Late-Life Cognitive Function

Background

Although the harmful effects of heavy and abusive drinking on health and lifespan have been unambiguously documented, less certain are the potential health benefits of moderate levels of alcohol consumption (Rehm, Gmel, Sempes, & Trevisan, 2003). There have been more than 100 epidemiological investigations of the association of moderate alcohol consumption with disease morbidity and total mortality, with a particular focus on middle-age and elderly populations (Andreasson, 2007). A recent meta-analysis showed that although risk of several diseases (e.g., esophageal cancer, liver cirrhosis, and chronic pancreatitis) increased monotonically with increasing level of alcohol consumption, risk of cardiovascular disease was lower among light and moderate drinkers than among heavy drinkers or those who were abstinent (Corrao, Bagnardi, Zambon, & La Vecchia, 2004).

Less clear is whether the presumed benefits of moderate drinking extend to noncardiovascular outcomes and especially cognitive outcomes. Although there is evidence that moderate levels of alcohol consumption may be protective against cognitive decline and the onset of dementia (Stampfer, Kang, Chen, Cherry, & Grodstein, 2005), other studies have failed to find an association between moderate alcohol consumption and cognitive functioning or change in cognitive functioning (Hebert et al., 1993). Shaper, Wannamethee, and Walker (1988) hypothesized that the apparent benefits of moderate drinking might be a consequence of inclusion in the alcohol abstinent group individuals who had quit drinking

due to the onset of illness or cognitive impairment. Fillmore, Kerr, Stockwell, Chikritzhs, and Bostrom (2006) recently published a meta-analysis showing that when individuals who reduced their drinking due to ill health are eliminated from the sample, the association of moderate drinking with health disappears. Fillmore's methodology and interpretations are controversial and have been criticized (Klatsky, 2007; Rehm, 2007), so the beneficial effects of moderate drinking remain uncertain.

Longitudinal Study of Aging Danish Twins (LSADT)

LSADT began in 1995 with an assessment of surviving twins from same-sex twin pairs born in Denmark before 1920 and continued with successive assessments of the original cohort, as well as subsequent birth cohorts who had aged into the catchment age range every 2 years through 2005 (K. Christensen, Holm, McGue, Corder, & Vaupel, 1999). A total of over 4,700 individuals and over 1,100 intact twin pairs age 70 years and older have participated in at least one wave of the six-wave longitudinal study. As part of their intake interview assessment, participants reported whether they drank alcohol and, if they did, how many drinks they had in a typical week. Responses to these questions were used to classify participants as either current light-to-moderate drinkers or nondrinkers (the questions did not allow us to differentiate heavy drinkers from light-to-moderate drinkers). The sample used in the present analysis consisted of 412 MZ (23 concordant nondrinkers, 309 concordant drinkers, and 80 discordant) and 597 same-sex DZ (29 concordant nondrinkers, 407 concordant drinkers, and 161 discordant) pairs. Rate of nondrinking did not differ significantly in MZ or DZ pairs (15.3% vs. 17.1%), but MZ pairs were more similar in their drinking status (tetrachoric correlation, $r_t = .46$) than DZ pairs ($r_t = .20$), indicating that drinking status is heritable and raising the possibility that any association of drinking with outcome may be due to genetic confounding. Cognitive functioning was assessed with the *Mini Mental State Examination* (MMSE; Folstein, Folstein, & McHugh, 1975) and with a composite of five brief cognitive tests selected to be sensitive to age-related changes (McGue & Christensen, 2001). To facilitate comparisons, researchers scaled both the cognitive composite and the MMSE to a T score metric (mean of 50 and standard deviation of 10). At the individual level, alcohol use was associated with a moderate but highly significant mean increase in performance on both the cognitive composite (mean T score difference of 3.2 ± 0.5 , $p < .0001$) and the MMSE (mean T score difference of 3.4 ± 0.5 , $p < .0001$).

Within- and Between-Pair Analyses

Begg and Parides (2003) and Gurrin, Carlin, Sterne, Dite, and Hopper (2006) provide a mixed-level regression framework for investigating within-pair and between-pair relationships of exposure with outcome. Let y_{ij} be the observed outcome for the j th twin ($j = 1, 2$) in the i th twin pair ($i = 1, 2, \dots, N$) and let x_{ij} be the corresponding exposure index for this individual. Then the overall, or individual-level, regression of outcome on exposure is given by the regression model

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \varepsilon_{ij}, \quad (1)$$

where β_1 is the individual-level effect of exposure (here alcohol consumption) on outcome (cognitive functioning), β_0 is the intercept term, and ε_{ij} is the residual (here, correlated across the two members of a twin pair). The overall regression effect can be further represented in terms of a within-pair (β_w) and a between-pair (β_B) effect using the regression model

$$y_{ij} = \beta_0 + \beta_w(x_{ij} - \bar{x}_i) + \beta_b\bar{x}_i + \varepsilon_{ij}, \quad (2)$$

where $\bar{x}_i$ is the mean exposure index for the i th twin pair. The within-pair regression coefficient provides a direct estimate of the effect of alcohol consumption within discordant-twin pairs.

We fit the models given in Equations 1 and 2 to the multiple waves of LSADT cognitive data using SAS Proc Mixed (Littell, Milliken, Stroup, & Wolfinger, 1996). Because twins within a pair are matched for age and gender, these demographic factors were treated as covariates in the between-pair analyses. The major results of these analyses are summarized in Figure 2. For both the cognitive composite and MMSE, moderate drinkers score about one third of a standard deviation higher on both cognitive measures (i.e., the individual-level effect). This drinking effect remained statistically significant within DZ pairs discordant for drinking but not within MZ discordant pairs. That is, within the 80 pairs of MZ twins who were discordant for drinking, the moderate-drinking twin scored on average 0.7 (± 0.9) T score units higher on the cognitive composite and 1.3 (± 1.0) points higher on the MMSE than his or her nondrinking cotwin, both nonsignificant differences.

Although the pattern of results is consistent with Scenario B in Figure 1, suggesting that genetic confounding rather than true causation accounts for the association of moderate drinking and cognitive functioning, it is important to evaluate these findings in light of the limitations discussed above. Given that we do not observe an association within MZ pairs, neither reverse causation (i.e., cognitive ability causing drinking) nor the contribution of unmeasured confounders underlying differences in exposure (e.g., nondrinkers are more likely than drinkers to be in poor health) seem to provide alternative explanations for our findings. Error in measuring alcohol exposure, however, could produce the pattern of results summarized in Figure 2. That is, since the drinking status of MZ twins is more highly correlated than the drinking status of DZ twins, measurement error could lead to greater attenuation of estimates within MZ than within DZ pairs. Unfortunately, we do not have an independent estimate of measurement error in our exposure index and so we cannot entirely rule out this possibility.

Summary

Developing effective prevention and intervention strategies will require that we understand the origins of what we are trying to prevent. Although experimental designs are rightly considered the most powerful approach for drawing causal inferences, they are not without limitations. Observational approaches have much to offer. The counterfactual or alternative outcomes model proposed by Rubin and colleagues (Rubin, 2007) provides a unified framework for evaluating causal inference and helps in identifying the relative strengths and weaknesses of both observational and experimental approaches. Experimental approaches may not always be feasible, and even when they are feasible, they may lead to biased estimation if only nonrepresentative subsamples of individuals are willing to submit to experimental assignment. The utility of observational approaches to causal inference will depend on the degree to which they lead to accurate estimation of the unobserved counterfactual: What would the outcomes have been among exposed individuals had they not been exposed? The essence to answering this question will depend on how observational researchers deal with confounding. Observational approaches based on statistically adjusting for confounders a posteriori although popular are often based on untested assumptions or fail to adequately consider the full range of confounders. Approaches that attempt to control for confounders through the use of observational research designs that approximate true

experiments have great potential in gerontology, even if they have been only infrequently employed and their limitations have not always been noted.

We focused on one natural experiment that might be of utility to psychological researchers: the discordant-twin design. MZ twin pairs who are discordant for exposure approximate, although clearly do not mirror, a true experiment. If exposure is causal, we expect that exposed members within MZ discordant pairs will show higher rates of the outcome than their unexposed cotwins. This is what was observed in the classic case of smoking and mortality. Unfortunately, there have been only a limited number of discordant-twin studies in gerontology. Findings from these studies suggest that social, intellectual, and physical activity may be protective against a range of impairments in late life. To take full advantage of the discordant-twin design, researchers will need large samples of twins to ensure observation of a sufficient number of discordant pairs and differential analysis of discordant MZ and DZ pairs to fully evaluate the genetic confounding alternative. Twin researchers need also be aware of the limitations of the discordant-twin design: (a) it does not necessarily rule out the possibility of reverse causation, (b) the factors that led to differences in exposure may account for differences in outcome, and (c) measurement error can result in differential attenuation of effect estimates by zygosity in within-pair analyses.

Acknowledgments

Funding

This work was supported by grants from the U.S. National Institute on Aging (P01-AG08761), National Institute on Alcohol Abuse and Alcoholism (R01 AA009367), and the Danish National Research Foundation.

References

- Andel R, Crowe M, Pedersen NL, Fratiglioni L, Johansson B, Gatz M. Physical exercise at midlife and risk of dementia three decades later: A population-based study of Swedish twins. *Journals of Gerontology: Series A Biological Sciences and Medical Sciences*. 2008; 63:62–66.
- Andel R, Crowe M, Pedersen NL, Mortimer J, Crimmins E, Johansson B, et al. Complexity of work and risk of Alzheimer's disease: A population-based study of Swedish twins. *Journals of Gerontology: Series B Psychological Sciences and Social Sciences*. 2005; 60:P251–P258.
- Anderson JC, Gerbing DW. Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*. 1988; 103:411–423.
- Andreasson S. The health benefits of moderate consumption called into question. *Addiction Research & Theory*. 2007; 15:3–5.
- Ashenfelter O, Krueger A. Estimates of economic return to schooling from a new sample of twins. *American Economic Review*. 1994; 84:1157–1173.
- Begg MD, Parides MK. Separation of individual-level and cluster-level covariate effects in regression analysis of correlated data. *Statistics in Medicine*. 2003; 22:2591–2602. [PubMed: 12898546]
- Bergen SE, Gardner CO, Aggen SH, Kendler KS. Socioeconomic status and social support following illicit drug use: Causal pathways or common liability? *Twin Research and Human Genetics*. 2008; 11:266–274. [PubMed: 18498205]
- Bentler, PM. *Theory and implementation of EQS: A structural equations program*. Los Angeles: BMDP Statistical Software; 1985.
- Bouchard, TJ.; Propping, P. *Twins as a tool of behavioral genetics*. Chichester, England: Wiley; 1993.
- Cacioppo JT, Hughes ME, Waite LJ, Hawkley LC, Thisted RA. Loneliness as a specific risk factor for depressive symptoms: Cross-sectional and longitudinal analyses. *Psychology and Aging*. 2006; 21:140–151. [PubMed: 16594799]
- Carlsson S, Andersson T, Lichtenstein P, Michaelsson K, Ahlbom A. Physical activity and mortality: Is the association explained by genetic selection? *American Journal of Epidemiology*. 2007; 166:255–259. [PubMed: 17493950]

- Carlsson S, Hammar N, Grill V, Kaprio J. Alcohol consumption and the incidence of type 2 diabetes: A 20-year follow-up of the Finnish Twin Cohort Study. *Diabetes Care*. 2003; 26:2785–2790. [PubMed: 14514580]
- Carmelli D, Page WF. Twenty-four year mortality in World War II U.S. male veteran twins discordant for cigarette smoking. *International Journal of Epidemiology*. 1997; 26:241–243. [PubMed: 9126527]
- Christensen H, Korten A, Jorm AF, Henderson AS, Scott R, Mackinnon AJ. Activity levels and cognitive functioning in an elderly community sample. *Age and Ageing*. 1996; 25:72–80. [PubMed: 8670533]
- Christensen K, Holm NV, McGue M, Corder L, Vaupel JW. A Danish population-based twin study on general health in the elderly. *Journal of Aging and Health*. 1999; 11:49–64. [PubMed: 10848141]
- Christensen K, Petersen I, Skytthe A, Herskind AM, McGue M, Bingley P. Comparison of academic performance of twins and singletons in adolescence: Follow-up study. *British Medical Journal*. 2006; 333:1095–1097. [PubMed: 17012267]
- Christensen K, Vaupel JW, Holm NV, Yashin AI. Mortality among twins after age-6: Fetal origins hypothesis versus twin method. *British Medical Journal*. 1995; 310:432–436. [PubMed: 7873948]
- Corrao G, Bagnardi V, Zambon A, La Vecchia C. A meta-analysis of alcohol consumption and the risk of 15 diseases. *Preventive Medicine*. 2004; 38:613–619. [PubMed: 15066364]
- Crowe M, Andel R, Pedersen N, Johansson B, Gatz M. Does participation in leisure activities lead to reduced risk of Alzheimer's disease? A prospective study of Swedish twins. *Journal of Gerontology: Psychological Sciences*. 2003; 58B:P249–P255.
- Depp CA, Jeste DV. Definitions and predictors of successful aging: A comprehensive review of larger quantitative studies. *American Journal of Geriatric Psychiatry*. 2006; 14:6–20. [PubMed: 16407577]
- Elzen H, Slaets JPJ, Snijders TAB, Steverink N. Do older patients who refuse to participate in a self-management intervention in the Netherlands differ from older patients who agree to participate? *Aging Clinical and Experimental Research*. 2008; 20:266–271. [PubMed: 18594195]
- Fillmore KM, Kerr WC, Stockwell T, Chikritzhs T, Bostrom A. Moderate alcohol use and reduced mortality risk: Systematic error in prospective studies. *Addiction Research & Theory*. 2006; 14:101–132.
- Finkel D, McGue M. Sex differences and nonadditivity in heritability of the Multidimensional Personality Questionnaire scales. *Journal of Personality and Social Psychology*. 1997; 72:929–938. [PubMed: 9108705]
- Fisher RA. Dangers of cigarette-smoking. *British Medical Journal*. 1957; 2:297–298.
- Fisher RA. Cancer and smoking. *Nature*. 1958; 182:596. [PubMed: 13577916]
- Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatry Research*. 1975; 12:189–198.
- Frederiksen H, Christensen K. The influence of genetic factors on physical functioning and exercise in second half of life. *Scandinavian Journal of Medicine and Science in Sports*. 2003; 13:9–18. [PubMed: 12535312]
- Freedman DA. Statistical models for causation: What inferential leverage do they provide? *Evaluation Review*. 2006; 30:691–713. [PubMed: 17093105]
- Glass TA, De Leon CFM, Bassuk SS, Berkman LF. Social engagement and depressive symptoms in late life: Longitudinal findings. *Journal of Aging and Health*. 2006; 18:604–628. [PubMed: 16835392]
- Griliches Z. Sibling models and data in economics: Beginnings of a survey. *Journal of Political Economy*. 1979; 87:S37–S64.
- Gurrin LC, Carlin JB, Sterne JAC, Dite GS, Hopper JL. Using bivariate models to understand between- and within-cluster regression coefficients, with application to twin data. *Biometrics*. 2006; 62:745–751. [PubMed: 16984316]
- Hebert LE, Scherr PA, Beckett LA, Albert MS, Rosner B, Taylor JO, et al. Relation of smoking and low-to-moderate alcohol-consumption to change in cognitive function: A longitudinal study in a defined community of older persons. *American Journal of Epidemiology*. 1993; 137:881–891. [PubMed: 8484379]

- Hernán MA. A definition of causal effect for epidemiological research. *Journal of Epidemiology and Community Health*. 2004; 58:265–271. [PubMed: 15026432]
- Hertzog C, Hultsch DF, Dixon RA. On the problem of detecting effects of lifestyle on cognitive change in adulthood: Reply to Pushkar et al. (1999). *Psychology and Aging*. 1999; 14:528–534.
- Johnson W, Krueger RF, Bouchard TJ, McGue M. The personalities of twins: Just ordinary folks. *Twin Research*. 2002; 5:125–131. [PubMed: 11931690]
- Kaprio J, Buchsbaum, MS.; Gottesman, II.; Heath, AC.; Körner, J.; Kringlen, E., et al. Group report: What can twin studies contribute to the understanding of adult psychopathology?. In: Bouchard, TJ.; Propping, P., editors. *Twins as a tool of behavioral genetics*. Chichester, England: Wiley; 1993. p. 287-299.
- Kaprio J, Koskenvuo M. Twins, smoking and mortality: A 12-year prospective study of smoking discordant twin pairs. *Social Science & Medicine*. 1989; 29:1083–1089. [PubMed: 2814591]
- Kendler KS, Pedersen NL, Farahmand BY, Persson PG. The treated incidence of psychotic and affective illness in twins compared with population expectation: A study in the Swedish Twin and Psychiatric Registries. *Psychological Medicine*. 1996; 26:1135–1144. [PubMed: 8931159]
- Klatsky AL. Errors in selection of “error-free” studies. *Addiction Research & Theory*. 2007; 15:8–16.
- Kujala UM, Kaprio J, Koskenvuo M. Modifiable risk factors as predictors of all-cause mortality: The roles of genetics and childhood environment. *American Journal of Epidemiology*. 2002; 156:985–993. [PubMed: 12446254]
- Levitt SD, List JA. What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*. 2007; 21:153–174.
- Li MD, Cheng R, Ma JZ, Swan GE. A meta-analysis of estimated genetic and environmental effects on smoking behavior in male and female adult twins. *Addiction*. 2003; 98:23–31. [PubMed: 12492752]
- Littell, RC.; Milliken, GA.; Stroup, WW.; Wolfinger, RD. *SAS system for mixed models*. Cary, NC: SAS Institute; 1996.
- Little RJ, Rubin DB. Causal effects in clinical and epidemiological studies via potential outcomes: Concepts and analytical approaches. *Annual Review of Public Health*. 2000; 21:121–145.
- Lynskey MT, Heath AC, Bucholz KK, Slutske WS, Madden PAF, Nelson EC, et al. Escalation of drug use in early-onset cannabis users vs. co-twin controls. *Journal of the American Medical Association*. 2003; 289:427–433. [PubMed: 12533121]
- Mackinnon A, Christensen H, Hofer SM, Korten AE, Jorm AF. Use it and still lose it? The association between activity and cognitive performance established using latent growth techniques in a community sample. *Aging Neuropsychology and Cognition*. 2003; 10:215–229.
- McGue M, Bouchard TJ. Genetic and environmental influences on human behavioral differences. *Annual Review of Neuroscience*. 1998; 21:1–24.
- McGue M, Christensen K. The heritability of cognitive functioning in very old adults: Evidence from Danish twins aged 75 years and older. *Psychology and Aging*. 2001; 16:272–280. [PubMed: 11405315]
- McGue M, Christensen K. Social activity and healthy aging: A study of aging Danish twins. *Twin Research and Human Genetics*. 2007; 10:255–265. [PubMed: 17564515]
- McNamee R. Confounding and confounders. *Occupational and Environmental Medicine*. 2003; 60:227–234. [PubMed: 12598677]
- Meehl PE. High school yearbooks: A reply to Schwarz. *Journal of Abnormal Psychology*. 1971; 77:143–148.
- Neyman, JS.; Dabrowska, DM.; Speed, TP. On the application of probability theory to agricultural experiments: Essay on principles; *Statistical Science*. 1990. p. 465-472. (Original work published 1923)
- Osler M, McGue M, Christensen K. Socioeconomic position and twins’ health: A life-course analysis of 1266 pairs of middle-aged Danish twins. *International Journal of Epidemiology*. 2007; 36:77–83. [PubMed: 17251245]
- Osler M, McGue M, Lund R, Christensen K. Marital status and twins’ health and behavior: An analysis of middle-aged Danish twins. *Psychosomatic Medicine*. 2008; 70:482–487. [PubMed: 18480194]

- Ozer DJ, Benet-Martinez V. Personality and the prediction of consequential outcomes. *Annual Review of Psychology*. 2006; 57:401–421.
- Parascandola M, Weed DL. Causation in epidemiology. *Journal of Epidemiology and Community Health*. 2001; 55:905–912. [PubMed: 11707485]
- Pearl, J. *Causality: Models, reasoning, and inference*. Cambridge, England: Cambridge University Press; 2000.
- Plomin R, Bergeman CS. The nature of nurture: Genetic influence on environmental measures. *Behavioral and Brain Sciences*. 1991; 14:373–385.
- Potter GG, Helms MJ, Burke JR, Steffens DC, Plassman BL. Job demands and dementia risk among male twin pairs. *Alzheimers & Dementia*. 2007; 3:192–199.
- Pushkar D, Etezadi J, Andres D, Arbuckle T, Schwartzman AE, Chaikelson J. Models of intelligence in late life: Comment on Hultsch et al. (1999). *Psychology and Aging*. 1999; 14:520–527. [PubMed: 10509704]
- Rehm J. On the limitations of observational studies. *Addiction Research & Theory*. 2007; 15:20–22.
- Rehm J, Gmel G, Sempes CT, Trevisan M. Alcohol-related morbidity and mortality. *Alcohol Research & Health*. 2003; 27:39–51. [PubMed: 15301399]
- Rejeski WJ, Fielding RA, Blair SN, Guralnik JM, Gill TM, Hadley EC, et al. The Lifestyle Interventions and Independence for Elders (LIFE) pilot study: Design and methods. *Contemporary Clinical Trials*. 2005; 26:141–154. [PubMed: 15837437]
- Rogosa D. A critique of cross-lagged correlation. *Psychological Bulletin*. 1980; 88:245–258.
- Ross L, Thomsen BL, Boesen EH, Johansen C. In a randomized controlled trial, missing data led to biased results regarding anxiety. *Journal of Clinical Epidemiology*. 2004; 57:1131–1137. [PubMed: 15567628]
- Rowe JW, Kahn RL. Human aging: Usual and successful. *Science*. 1987; 237:143–149. [PubMed: 3299702]
- Rowe, JW.; Kahn, RL. *Successful aging*. New York: Dell; 1999.
- Rubin DB. Estimating the causal effects of smoking. *Statistics in Medicine*. 2001; 20:1395–1414. [PubMed: 11343361]
- Rubin DB. The design versus the analysis of observational studies for causal effects: Parallels with the design of randomized trials. *Statistics in Medicine*. 2007; 26:20–36. [PubMed: 17072897]
- Rubin DB. For objective causal inference, design trumps analysis. *Annals of Applied Statistics*. 2008; 2:808–840.
- Rutter M. Proceeding from observed correlation to causal inference. *Perspectives on Psychological Science*. 2007; 2:377–395.
- Salthouse TA. Mental exercise and mental aging: Evaluating the validity of the “use it or lose it” hypothesis. *Perspectives on Psychological Science*. 2006; 1:68–87.
- Scarr S, McCartney K. How people make their own environments: A theory of genotype-environment effects. *Child Development*. 1983; 54:424–435. [PubMed: 6683622]
- Shadish, WR.; Cook, TD.; Campbell, DT. *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin; 2002.
- Shaper AG, Wannamethee G, Walker M. Alcohol and mortality in British men: Explaining the U-shaped curve. *Lancet*. 1988; 2:1267–1273. [PubMed: 2904004]
- Stampfer MJ, Kang JH, Chen J, Cherry R, Grodstein F. Effects of moderate alcohol consumption on cognitive function in women. *New England Journal of Medicine*. 2005; 352:245–253. [PubMed: 15659724]
- Stubbe JH, de Moor MHM, Boomsma DI, de Geus EJC. The association between exercise participation and well-being: A co-twin study. *Preventive Medicine*. 2007; 44:148–152. [PubMed: 17059845]
- Tomarken AJ, Waller NG. Structural equation modeling: Strengths, limitations, and misconceptions. *Annual Review of Clinical Psychology*. 2005; 1:31–65.
- Trost SG, Owen N, Bauman AE, Sallis JF, Brown W. Correlates of adults’ participation in physical activity: Review and update. *Medicine and Science in Sports and Exercise*. 2002; 34:1996–2001. [PubMed: 12471307]

- van Heuvelen MJG, Hochstenbach JBM, Brouwer WH, de Greef MHG, Zijlstra GAR, van Jaarsveld E, et al. Differences between participants and non-participants in an RCT on physical activity and psychological interventions for older persons. *Aging Clinical and Experimental Research*. 2005; 17:236–245. [PubMed: 16110738]
- Vink JM, Nawijn L, Boomsma DI, Willemsen G. Personality differences in monozygotic twins discordant for cannabis use. *Addiction*. 2007; 102:1942–1946. [PubMed: 18031429]
- Waller K, Kaprio J, Kujala UM. Associations between long-term physical activity, waist circumference and weight gain: A 30-year longitudinal twin study. *International Journal of Obesity*. 2008; 32:353–361. [PubMed: 17653065]
- Weinberg CR. Toward a clearer definition of confounding. *American Journal of Epidemiology*. 1993; 137:1–8. [PubMed: 8434568]
- West SG, Duan N, Pequegnat W, Gaist P, Jarlais DCD, Holtgrave D, et al. Alternatives to the randomized controlled trial. *American Journal of Public Health*. 2008; 98:1359–1366. [PubMed: 18556609]
- Whitfield JB, Zhu G, Madden PA, Neale MC, Heath AC, Martin NG. The genetics of alcohol intake and of alcohol dependence. *Alcoholism: Clinical and Experimental Research*. 2004; 28:1153–1160.

Fig. 1.

Logic of the discordant-twin design. We graphed the hypothetical effect of exposure on outcome when measured at the individual level (IL; i.e., without regard to twin-pair membership) and within monozygotic (MZ) and dizygotic (DZ) twin pairs discordant on exposure. Scenario A is consistent with an association between exposure and outcome that is due entirely to a causal effect. Scenario B depicts an association that is due entirely to confounding: The within-MZ pair effect is zero and thus does not provide evidence in support of a causal effect. Scenario C depicts an association that is due in part to confounding, but the existence of an effect within discordant MZ twins is consistent with some causal influence.

Fig. 2.

Analysis of the effect of drinking (exposure) on cognitive functioning (outcome) in the Longitudinal Study of Aging Danish Twins (LSADT). We plotted the estimated mean difference between moderate drinkers and abstainers at the individual level, as well as within monozygotic (MZ) and dizygotic (DZ) twin pairs discordant for drinking. Cognitive functioning was assessed using both a five-test cognitive composite and the Mini Mental State Examination (MMSE). To facilitate comparison across outcomes, each outcome was scaled to a *T* score metric (i.e., mean of 50 and standard deviation of 10) at the intake assessment. Error bars demarcate one standard error. Moderate drinking was associated with significantly increased cognitive performance in individual-level analyses and within discordant DZ twin pairs. A significant effect was not observed within discordant MZ pairs.

Table 1

Discordant-Twin Studies in Social and Psychological Gerontology

Study and outcome	Exposure	Is there a significant association between exposure and outcome?			
		Individual level (N)	Within twin pairs (N)	Within MZ pairs (N)	Within DZ pairs (N)
Andel et al. (2005)					
Dementia	Job complexity	Yes (10,079)	Yes (55)	NR	NR
Andel et al. (2008)					
Dementia	Physical exercise	Yes (3,134)	No (90)	NR	NR
Carlsson et al. (2003)					
Type 2 diabetes	Alcohol consumption	Yes (22,778)	NR	NR	No (27)
Carlsson et al. (2007)					
Mortality	Physical activity	Yes (26,218)	Yes (8,696)	No (3,112)	No (5,584)
Crowe et al. (2003)					
Dementia	Leisure activity	NR	Yes (107)	NR	NR
Kujala et al. (2002)					
Mortality	Exercise	Yes (15,904)	No (658)	No (155)	No (475)
Mortality	Smoking	Yes (15,904)	Yes (809)	Yes (166)	Yes (591)
Mortality	Heavy alcohol	Yes (15,904)	Yes (507)	Yes (105)	Yes (376)
McGue et al. (2007)					
Physical function	Social activity	Yes (4,731)	NR	Yes (70)	NR
Cognitive	Social activity	Yes (4,731)	NR	Yes (70)	NR
Depression	Social activity	Yes (4,731)	NR	Yes (70)	NR
Osler et al. (2007) ^a					
Cognitive	Social class	Yes (2,441)	NR	No (NR)	Yes (NR)
BMI	Social class	Yes (2,420)	NR	No (NR)	No (NR)
Self-rated health	Social class	Yes (2,441)	NR	No (NR)	No (NR)
Osler et al. (2008)					
Cognitive	Married	Yes (2,270)	Yes (333)	No (NR)	Yes (NR)
Smoking	Married	Yes (2,274)	No (333)	No (NR)	No (NR)
Depression	Married	Yes (2,270)	Yes (333)	Yes (NR)	Yes (NR)
Potter et al. (2007)					

Study and outcome	Exposure	Is there a significant association between exposure and outcome?			
		Individual level (N)	Within twin pairs (N)	Within MZ pairs (N)	Within DZ pairs (N)
Dementia	Job complexity	Yes (425)	Yes (220)	NR	NR
Stubbe et al. (2007)					
Life satisfaction	Physical exercise	Yes (8,000)	NR	No (174)	No (162)
Waller et al. (2008)					
30-year weight gain	Physical activity	Yes (178)	Yes (42)	NR	NR

Note. NR = Result or sample size not reported.

^aIn the Osler et al. (2007) study only the total number of discordant twin pairs was reported ($N = 386$).