Journal of the International Neuropsychological Society (2009), **15**, 331–343.
Copyright © 2009 INS. Published by Cambridge University Press. Printed in the USA.
doi:10.1017/S1355617709090481

CRITICAL REVIEW

Why IQ is not a covariate in cognitive studies of neurodevelopmental disorders

MAUREEN DENNIS,[1,2] DAVID J. FRANCIS,[3,4] PAUL T. CIRINO,[3,4] RUSSELL SCHACHAR,[1,5] MARCIA A. BARNES,[6] AND JACK M. FLETCHER[3]

[1]Program in Neurosciences and Mental Health, Department of Psychology, The Hospital for Sick Children, Toronto, Ontario, Canada
[2]Department of Surgery, The Hospital for Sick Children, Toronto, Ontario, Canada
[3]Department of Psychology, University of Houston, Houston, Texas
[4]Texas Institute for Measurement, Evaluation, and Statistics, University of Houston, Houston, Texas
[5]Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada
[6]Children's Learning Institute, Department of Pediatrics, University of Texas Houston, Houston, Texas

(Received September 15, 2008; Final Revision December 29, 2008; Accepted December 29, 2008)

Abstract

IQ scores are volatile indices of global functional outcome, the final common path of an individual's genes, biology, cognition, education, and experiences. In studying neurocognitive outcomes in children with neurodevelopmental disorders, it is commonly assumed that IQ can and should be partialed out of statistical relations or used as a covariate for specific measures of cognitive outcome. We propose that it is misguided and generally unjustified to attempt to control for IQ differences by matching procedures or, more commonly, by using IQ scores as covariates. We offer logical, statistical, and methodological arguments, with examples from three neurodevelopmental disorders (spina bifida meningomyelocele, learning disabilities, and attention deficit hyperactivity disorder) that: (1) a historical reification of general intelligence, *g*, as a causal construct that measures aptitude and potential rather than achievement and performance has fostered the idea that IQ has special status and that in studying neurocognitive function in neurodevelopmental disorders; (2) IQ does not meet the requirements for a covariate; and (3) using IQ as a matching variable or covariate has produced overcorrected, anomalous, and counterintuitive findings about neurocognitive function. (*JINS*, 2009, *15*, 331–343.)

Keywords: Intelligence, History, Intelligence tests, Meningomyelocele, Analysis of variance, Learning disability, Attention deficit disorder with hyperactivity

INTRODUCTION

Throughout the life span, IQ is a volatile index of global functional outcome, the final common path of an individual's genes, biology, cognition, education, and experiences. Studies of adult brain disorders are conducted largely without reference to IQ scores. For example, studies of adult aphasia ignore verbal IQ, even though it has long been recognized (Hebb, 1949) that the same brain injury that causes aphasia also disrupts intelligence. We are not aware of an adult outcome paper that treats postinjury IQ as a factor to be covaried out of postinjury measures of function.

Neurodevelopmental disorders occur early in development as a result of a congenital insult associated with altered genes and brains. Some are diagnosed on the basis of genetic and brain defects [e.g., spina bifida meningomyelocele (SBM) or Williams syndrome]. Others are identified by cognitive-behavioral deficits, which are typically accompanied by genetic and brain anomalies [e.g., learning disabilities (LD) or attention deficit hyperactivity disorder (ADHD)]. Neurodevelopmental disorders are different from adult acquired disorders [and from childhood acquired disorders involving traumatic brain injury (TBI), strokes, or tumors] in an important way: they involve no period of normal development.

Any IQ score in a neurodevelopmental disorder postdates (not predates) the condition, charts the history of the condition, is always confounded with and/or by the condition, and can never be separated from the effects of the condition.

Correspondence and reprint requests to: Maureen Dennis, Program in Neurosciences and Mental Health, Department of Psychology, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, Canada M5G 1X8. E-mail: maureen.dennis@sickkids.ca

https://doi.org/10.1017/S1355617709090481 Published online by Cambridge University Press

Nevertheless, it is not unusual for reviewers of neurodevelopmental studies to request that groups be matched/equated/controlled for IQ, with a common statistical recommendation being to covary IQ from specific cognitive measures.

The different treatments of IQ in neurodevelopmental disorders and adult acquired brain insults might suggest that intelligence is a construct to be treated separately from cognition only after an individual can drink or vote. We have resisted exploring this idea, beyond noting that it is incompatible with current views of neurocognitive development, which stress life span continuities as well as discontinuities (e.g., Craik & Bialystok, 2006). What does concern us is the use of IQ in an explanatory framework whereby general ability factors cause, and can therefore be separated from, more specific cognitive skills. In this article, we argue that it is misguided and generally unjustified to attempt to control for IQ differences by matching procedures or, more commonly, by using IQ scores as covariates, and we support the argument with specific examples from three neurodevelopmental disorders (SBM, LD, and ADHD) that (1) the special but spurious status of IQ as the generic covariate arose from a historical reification of general intelligence, *g*, as a causal construct that measures aptitude and potential, rather than achievement and performance and that in studying neurocognitive function in neurodevelopmental disorders; (2) IQ does not fulfill the methodological and statistical requirements of a covariate; and (3) the use of IQ as a matching variable or covariate has produced anomalous, overcorrected, counterintuitive, and theoretically vacuous findings about neurocognitive function.

THE HISTORICAL REIFICATION OF GENERAL INTELLIGENCE, *g*

To provide the groundwork for the statistical and methodological arguments that follow, we first consider the genesis of *g*, the *sine qua non* of IQ and its chief "active ingredient" (Jensen, 1989). As a general ability factor, *g* has come to represent a latent construct: people have more or less of *g*, and *g* measures their aptitude and potential, rather than their achievement and performance.

The Reification of *g*

The father of IQ testing, Alfred Binet (1857–1911) conceived of intelligence as a shifting complex of environmentally malleable, developmentally variable, and diverse functions (Binet & Simon, 1916; Evans & Waites, 1981; Siegler, 1992; Wolf, 1973), which was the basis for an ordinal ranking of performance rather than an absolute measure of capacity (Binet & Simon, 1916). Early in the history of intelligence testing in Britain and the United States, IQ became reified (Gould, 1981). The idea that IQ was a latent construct, not simply the sample or long-run average of a set of test scores, became quite pervasive with Terman's English language revision of the Binet–Simon tests (Terman, 1916) and, particularly, with Spearman's introduction of the construct of *g* (Spearman, 1904).

Spearman noted that correlations between pairs of tests form a "positive manifold" in which some portion of the variance in each test could be attributed to a universal general factor, *g*, common to all intelligent activities (Spearman, 1927). He considered that *g* was the "one great common Intellective Function" (Spearman, 1904, p. 51) and that all examinations of sensory, academic, or specific intellectual functions were independent estimates of *g*.

Despite early psychometric evidence against *g* [Thomson (1916, 1919) showed that intercorrelations among tests could produce hierarchies without invoking a general factor, so that *g* was extraordinarily improbable], disciples of *g* (e.g., Jensen, 1969) have argued that it has stood like "a rock of Gibraltar" and they have even presupposed its existence: ["… almost any *g* is a 'good' *g* and is certainly better than no *g*" (Jensen & Weng, 1994, p. 231)]. Some later intelligence theories have continued to embrace *g* (e.g., Vernon, 1964), while others have rejected it in favor of fluid and crystallized intelligence (Cattell, 1943; Horn, 1998). Others have included *g* within mental strata (Carroll, 1993) or tested its role in competing psychometric models of intelligence (e.g., Johnson & Bouchard, 2005). A persisting idea is that IQ is an entity, a latent variable in the strong "true score" sense of the term (Lord, 1965); for example, recent formulations of *g* highlight its content-free character, which allows an individual to deal with complexity and change (e.g., Lubinski, 2004).

Causal Hierarchies of *g*

In France, Binet's test had a relatively narrow application for individual academic diagnosis and the study of individual differences, with the goal of intelligence testing being to sketch the characteristic profile of individuals, not to establish a global hierarchy of intelligence (Piéron, 1932). From around 1914 and onward, *g* became associated with a number of social and political values (involving civic worth, eugenics, selective breeding, and immigration policy) in Britain (Evans & Waites, 1981), the United States (Kamin, 1974), and Europe and Scandinavia (Roll-Hansen, 1988). IQ testing became part of a large-scale evaluation system concerned with ranking large numbers of individuals in a hierarchy based on social class or race (Schneider, 1992). High *g* on Terman's Stanford–Binet IQ test became conflated with health, masculinity, and heterosexuality (Hegarty, 2007). IQ scores discriminated immigrant groups (Goddard, 1917), and later, *g* was suggested as the major systematic source of Black–White population differences (Jensen, 1985). In an ongoing interaction between scientific knowledge and political ideology (Roll-Hansen, 1988), research findings about IQ have often been assessed not so much on their scientific standing as on their supposed political implications (Neisser et al., 1996). The idea that *g* causes individual and group differences remains current, with recent arguments that *g* is the underlying cause of health inequalities among socioeconomic groups (Gottfredson, 2004).

https://doi.org/10.1017/S1355617709090481 Published online by Cambridge University Press

The Invariant *g* Argument

Spearman argued that the proof of *g* was independent of test conditions, test procedures, test reliability, homogeneity of the group of people being tested, historical times, geographies, and cultures; he described *g* as "reproducible at all times, places and manners …" (Spearman, 1904, p. 50). However, definitions and measures of intelligence appear to be shaped by time, place, culture (Kornhaber et al., 1990), and brains.

IQ scores have changed over historical time, both rising (Flynn, 2007) and falling (Teasdale & Owen, 2008); furthermore, when IQ scores rise with successive standardization samples, the "*g*-ness" of the tests (their average intercorrelation) falls, particularly for Performance IQ (Kane & Oakland, 2000). IQ also varies with intracontinental geography at the same historical time; Goodenough (1949, table 2, p. 17) showed that 21.7% of girls in Birmingham, Alabama, but only 5.5% of girls in Los Angeles, California, were three grades delayed academically. The assertion that Australian aborigines had lower intelligence on Porteus's "culture-free" pencil-and-paper mazes measure (Porteus, 1917) neglected to consider that the essential feature of his maze test, the *cul-de-sac*, does not exist in the featureless 1.3 million square miles of the Great Australian Desert (Lynn, 1978).

Spearman conceived of *g* as a marker for innate mental energic capacity (Spearman, 1914), a view that, according to Evans and Waites (1981), he supported by appeals to contemporary neurophysiology. However, there is no single brain location for *g*; brain lesions do not disrupt outcome in proportion to the *g* loading of the IQ task; different brain configurations are consistent with equivalent IQ scores (Haier et al., 2005); the relative contributions of gray and white matter to explaining variations in IQ shifts with age (Haier et al., 2004; Johnson et al., 2008; Jung & Haier, 2007), and the strong correlation between whole-brain gray matter volume and IQ develops only gradually (Wilke et al., 2003).

As we argue next, even if we agree that *g* is real, that IQ tests measure *g* independent of how it is assessed, or that IQ is sufficiently invariant and stable to measure core capacity, it cannot be controlled statistically, and covariance analyses do not eliminate *g* or IQ as the cause of specific cognitive outcomes.

IQ AS A COVARIATE: METHODOLOGICAL AND STATISTICAL ISSUES

Methodological Issues

Analysis of covariance (ANCOVA) was devised for classical experimental designs with random group assignment to minimize preexisting group differences, a situation where group differences in characteristics like IQ or socioeconomic status (SES) occur only by chance, the theoretical populations to which the experimenter wishes to generalize being equated on the distribution of the covariate. Even with random assignment, study differences may occur on the covariate by chance, so ANCOVA is a possible means of adjusting for sample differences on the covariate and providing an unbiased estimate of the population difference in means on the dependent variable (because the hypothetical populations to which the treatments have been assigned have been equated by design).

When the covariate is an attribute of the disorder or of its treatment, or is intrinsic to the condition, it becomes meaningless to "adjust" the treatment effects for differences in the covariate, and ANCOVA cannot be used to control treatment assignment independent of the covariate (Adams et al., 1985; Evans & Anastasio, 1968; Lord, 1967, 1969; Miller & Chapman, 2001; Tupper & Rosenblood, 1984). In his classic demonstration of an agronomist comparing rates of growth in corn plants that differ inherently in stalk height, Lord (1969) showed that any attempt to compare the yields of the two classes of plants by adjusting for plant height must give a meaningless result, one that could only come about through fundamental alterations of the two plants. The causal network relating plant species to plant height and plant yield cannot be manipulated to isolate the causal impact of species on yield in the absence of species effects on height and height effects on yield; neither ANCOVA nor matching can correct these effects of species and height.

The best case scenario for the use of a covariate (Huitema, 1980) exists when: (a) the assignment to the independent variable (e.g., neurodevelopmental disorder) is done randomly; (b) the covariate is related to the outcome measure, but this relation is of no theoretical interest in terms of the investigative question (i.e., the covariate is a source of irrelevant variation in the dependent variable, which, if controlled, allows for a more powerful test of the effects of the independent variable of interest]; (c) the covariate is unrelated to the independent variable, which is assured probabilistically if (a) is true; and (d) the covariate is not differentially related to the dependent variable at different levels of the independent variable [also assured if (a) is true]. Ideally, the covariate should also be stable and measured without error.

When assignment to the independent variable is not through randomization, or the covariate otherwise does not meet all the requirements of the ideal scenario, then their proper use requires consideration of precisely how the independent variable, the dependent variable, and the covariate come together to form a causal network. For instance, covariates can meaningfully be incorporated into the analysis when the dependent and independent variables are spuriously related to the covariate, or when the covariate mediates (partially or fully) the relation between the independent and the dependent variable, and the investigator is interested in estimating the direct effect of the independent variable on the outcome. In these instances, the use of a covariate can clarify the relation between the independent and the dependent variables.

We next argue that the typical use of IQ as a covariate does not fulfill the requirements of the ideal scenario. Furthermore, it rarely meets the requirements for the meaningful use of covariates in less than ideal circumstances.

https://doi.org/10.1017/S1355617709090481 Published online by Cambridge University Press

The Ideal Scenario for a Covariate

At the heart of an ideal scenario and all meaningful uses of covariates is the tripartite relation of the covariate, the independent variable, and the outcome. In appropriate uses of covariates, the covariate is a cause of the outcome, such as age causing achievement, or at least serving as a proxy for exposure, education, and instruction. The covariate should not be an outcome of the dependent variable or of the independent variable. In this three-dimensional space, what complicates matters is the relation between the covariate and the independent variable, and by implication, the joint relations among the independent variable, the dependent variable, and the covariate. When assignment to values on the independent variable is through a random process (the ideal scenario), the independent variable and the covariate are unrelated (i.e., the extent to which the groups differ on the covariate is probabilistically zero), and the inclusion of covariates in the statistical analysis increases power for finding a true relation between the independent and the dependent variables by keeping the numerator of the F value the same while reducing the denominator.

This situation is depicted graphically in Figure 1. Although the situation depicted in Figure 1 is hypothetical, we have labeled the horizontal axis as IQ and the vertical axis as Memory to make the situation less abstract. In Figure 1, the difference in the heights of the two ellipses at the mean of IQ is equivalent to the difference between the groups' means on the Memory measure. The difference in this adjusted comparison is not in the estimate of the mean difference between groups, but in terms of the variance in Memory. In the comparison of Memory controlling for IQ, the variance of Memory is replaced by the variance in Memory conditional on IQ. Given the correlation of .6 in the population, the conditional variance in Memory will be about 64% of the unconditional variance in Memory, thereby leading to a more powerful test of the difference between groups on Memory.

The Less-Than-Ideal Scenario for a Covariate

When preexisting groups are compared in a nonexperimental study, participants are recruited nonrandomly, as they exist in nature. If we knew how children come to be "assigned" to the population of children with SBM or LD, it might be possible to incorporate the assignment process into the comparison; even for genetic disorders, however, modeling the selection process is not currently possible, so groups may differ on variables potentially related to the assignment mechanism. It is a false inference that any measure on which groups differ and which is not itself the comparison of interest must be controlled because it is related to the assignment mechanism.

Many differences between naturally occurring groups are themselves consequences of the unknown assignment mechanism, being neither artifacts of how the relevant sample was ascertained nor part of the assignment mechanism, but rather differences between the populations from which the researcher wishes to sample. Investigators understandably wish to adjust for selection effects that arise due to nonrepresentative sampling from the populations, in order to derive a better estimate of population differences by adjusting for sampling biases, such as differences in age or gender. But when the populations differ on the attribute, even random sampling from the populations will result in attribute differences between samples that represent not biased sampling but true population differences.

For groups with neurodevelopmental disorders, mean IQ scores will be generally below the population normative mean. Consequently, groups will differ when appropriately selected from the populations of these disorders. Differences in IQ between children with SB and age-matched controls represent, not poor sampling, but preexisting, nonrandom differences beyond experimenter control.

This situation is depicted in Figure 2, which is developed in a fashion similar to Figure 1, but allows for differences between the two groups on the variable IQ. The distance between the two solid horizontal lines depicts the difference in Memory controlling for the differences between groups on IQ. Figure 2 depicts two population distributions, with the two distributions being closer together at the grand mean of IQ than at the respective group means on IQ. The two distributions are almost nonoverlapping, such that much less than 50% of the lower performing group lies at or above the grand mean on IQ, while substantially more than 50% of the higher performing group lies above the grand mean on IQ. In the hypothetical situation depicted in Figure 2, a comparison at the grand mean is roughly equivalent to comparing the 25th percentile in the higher performing group and the 75th percentile in the lower performing group. That this statistical adjustment can be performed mathematically says nothing about the scientific validity of the resulting comparison, which requires a model of the neurocognitive function.

The inability to control group assignment renders the foregoing discussion somewhat academic, insofar as it relates to controlling for preexisting differences on covariates. It does highlight the fact that the key to appropriate use of covariates is understanding their role in the assignment mechanism and the selection process and articulating a causal network about how different cognitive and neurodevelopmental processes are related.

Assumptions of ANCOVA

The use of IQ as a covariate in neurodevelopmental studies rarely meets standard assumptions for ANCOVA. In addition to the assumptions of analysis of variance, ANCOVA adds the assumption of homogeneity of regression, which practically means that the within-group regressions of IQ and the dependent variable are not different. ANCOVA assumes further that the residuals are normally distributed and have equal variance in all groups.

Although these assumptions can be relaxed with appropriate alternative estimation methods, consider what happens

https://doi.org/10.1017/S1355617709090481 Published online by Cambridge University Press

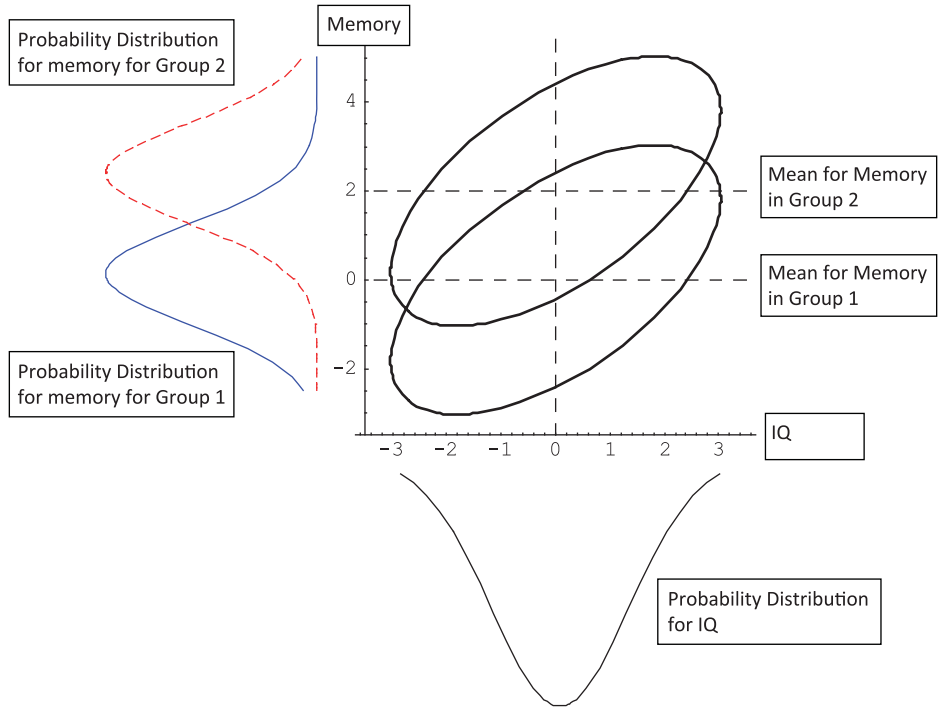

Fig. 1. The ellipses in the figure represent the 99% quantiles in a bivariate normal distribution for two groups where the correlation between IQ (graphed on the horizontal axis) and Memory (graphed on the vertical axis) is .6 for each group. In the margins of the figure are graphed the univariate probability density functions (i.e., the univariate distributions) for the groups. The normal distribution below the horizontal axis shows the marginal distribution for IQ, while the two normal distributions on the left side of the figure show the marginal distribution of Memory for Groups 1 and 2, respectively. Note that in the margin of the horizontal axis, only a single normal distribution is graphed because the populations are equated on IQ. In contrast, on the vertical axis, two distributions are shown in the margin, reflecting the difference in the mean of Memory for the two populations. The dashed horizontal lines are plotted at the mean of Memory for each group to make it easy to compare the group means on Memory. The vertical line is plotted at the mean of IQ. The fact that there is only one vertical line indicates that the populations are equated on IQ. If we ignore information about IQ, comparing the two groups on Memory would amount to examining the difference in the heights of the two horizontal lines relative to the variability in the marginal distributions of Memory (i.e., the normal distributions on the left side of the figure). When information on IQ is included in the analysis, then groups are implicitly being compared on Memory at the grand mean for IQ, and this difference is evaluated relative to the variability in the conditional distribution of Memory, which is much less than the variability in the marginal distribution of Memory.

when the covariate seemingly has no effect on the outcome or, conversely, when the covariate relates to the dependent variable in a different manner for each group, such that group differences in the outcome vary as a function of the value of the covariate. In the former situation, the lack of direct impact of the covariate on the dependent variable when the ANCOVA assumptions are met implies that the covariate does not mediate or moderate the relationship between the group measure and the dependent variable; such an inference is not necessarily justified if the assumptions of the ANCOVA model do not hold. The presence of a relation between the covariate and the dependent variable does not imply that the covariate mediates or moderates the relationship between the group measure and the dependent variable; such an inference requires a line of causal argument that is not simply statistical in nature, and so must be supported through both theory and empirical findings.

The alteration of group differences by inclusion of a covariate occurs when groups differ on the covariate or when the covariate operates differently in predicting group outcome; of itself, the alteration does not license the inference that the covariate mediates or moderates the relationship between the group and the dependent variable. In the absence of heterogeneity of regression, an adjustment in the mean difference occurs because the groups differ on average on the covariate, as shown in Figure 2. Comparing groups at the mean value of the covariate leads to a different estimate of group differences on the outcome than simply comparing groups on their unadjusted means on the dependent variable.

Controlling for the covariate usually reduces the magnitude of group differences, as shown in Figure 2, although this adjustment need not shrink group differences. In fact, when the covariate is positively related to the outcome within groups, but the lower scoring group is higher on the covariate, the adjusted mean difference will exceed the unadjusted mean difference in magnitude. This scenario is depicted in Figure 3, where there is homogeneity of regression but where group differences are larger when the covariate, IQ, is

https://doi.org/10.1017/S1355617709090481 Published online by Cambridge University Press

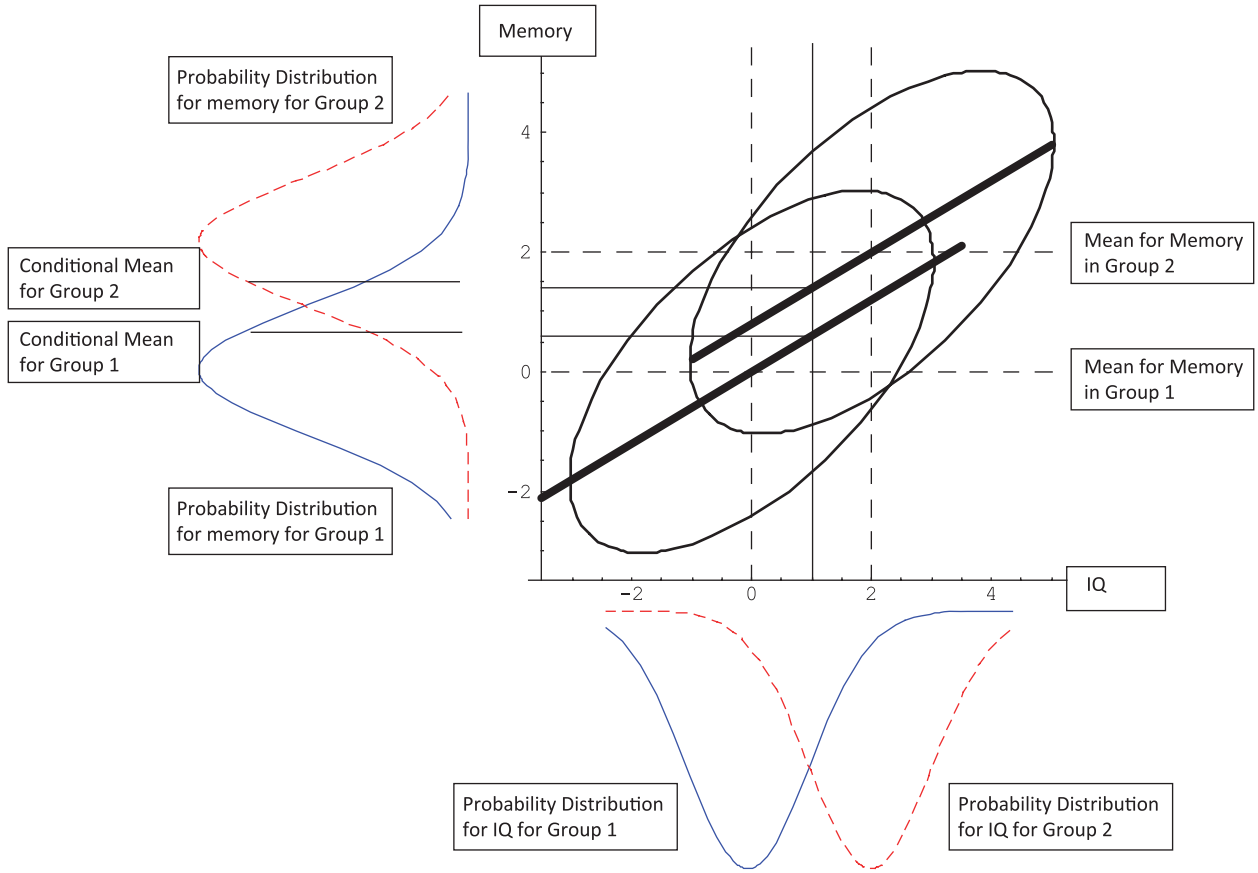

Fig. 2. This figure differs systematically from Figure 1 because the two groups differ on the mean of IQ. As in Figure 1, the correlation between the IQ and the Memory is .6 for each group. In addition, Figure 2 includes two heavy lines that depict the regression of Memory on IQ for each group and also includes a second set of horizontal lines. As in Figure 1, the horizontal dashed lines depict the unconditional mean of Memory for each group. The solid horizontal lines, in contrast, depict the conditional mean of Memory for each group; that is, the solid horizontal lines show the expected value for Memory, for individuals in each group with scores on IQ that are equal to the grand mean of IQ, which is depicted by the solid vertical line.

controlled than when the groups are compared on Memory ignoring IQ. This effect can be seen in Figure 3 by comparing the separation between the two solid horizontal lines, which show the difference in the adjusted means, to the separation between the two dashed horizontal lines, which show the difference in the unadjusted means, and which are also referenced by the centers of the marginal distributions for Memory. Such findings are possible when the covariate is causally implicated in the dependent variable but other factors operate to bias group selection.

When the relation between the covariate and the outcome is different for each of the two groups, differences on the outcome vary with the value of the covariate. In Figure 4, the ellipses are of different sizes reflecting the overall weaker relation between IQ and Memory in the lower performing group ($r = .4$ *vs.* $r = .8$ in the higher performing group). Differences between groups on the outcome measure Memory depends on where along the IQ distribution the comparison between groups is made. The standard ANCOVA comparison is made at the grand mean on IQ, which in this case represents approximately the 25th percentile for the higher performing group and the 75th percentile for the lower performing group. If a common regression line was applied to the two groups, the adjustment would be too little for the higher performing group (where dependence of Memory on IQ is greater) and too great for the lower performing group (where IQ and Memory are less strongly related).

ANOMALOUS RESULTS WHEN IQ IS USED AS A COVARIATE

Notwithstanding the logical and statistical issues discussed above, it is extremely common in studies of neurodevelopmental disorders to match for IQ or to use IQ as a covariate. We next consider the use of IQ as a measure of aptitude rather than achievement, discrepancy definitions of IQ, and how the use of IQ as a covariate shapes anomalous interpretations of outcome measures of neurodevelopmental disorders.

IQ as a Measure of Aptitude

Binet thought that intelligence was a crop whose yield could be enhanced with education. "… these deplorable

https://doi.org/10.1017/S1355617709090481 Published online by Cambridge University Press

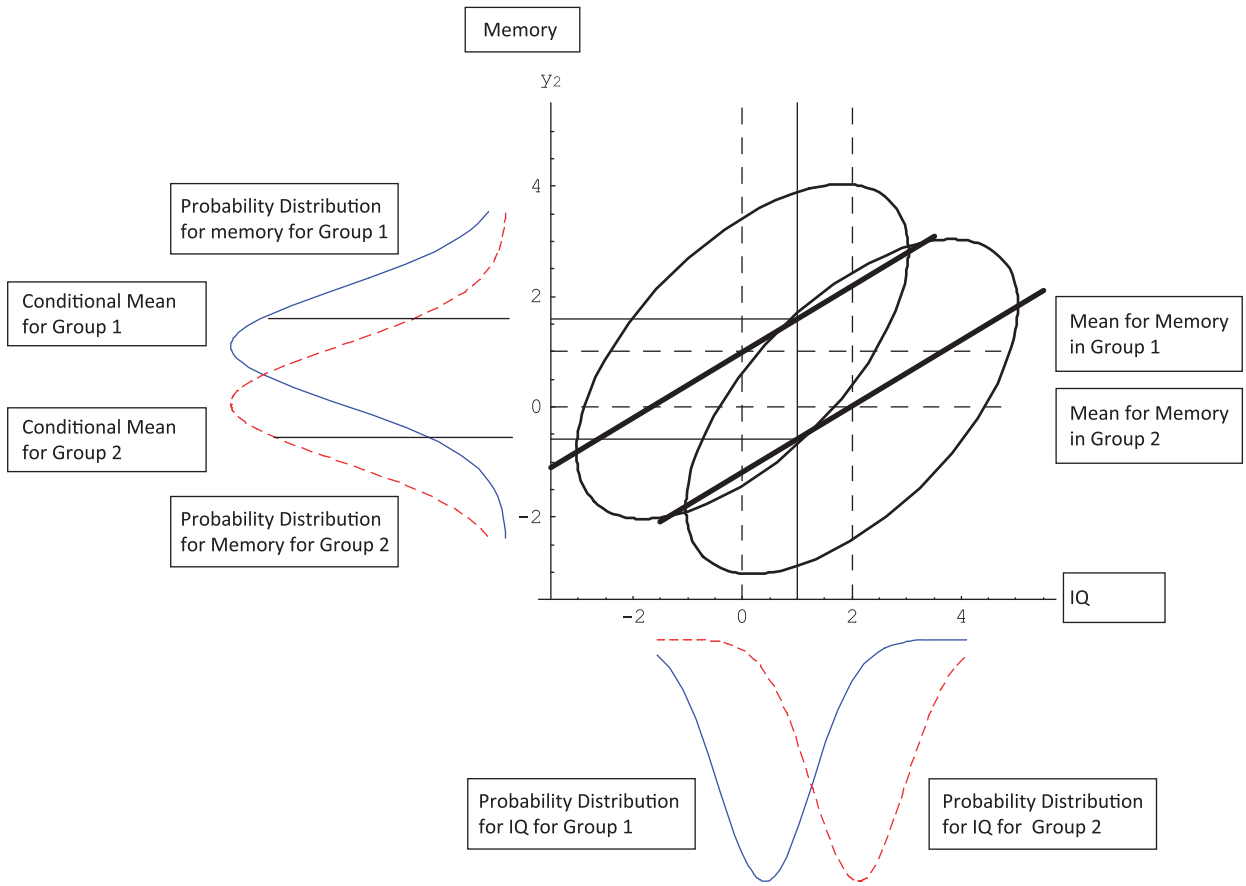

Fig. 3. This figure is similar to Figure 2, but in this case, the displacement of groups on IQ is opposite what would be expected given the overall positive correlation between IQ and Memory in both groups. The two solid horizontal lines show the difference in the conditional (i.e., adjusted) means. The separation between the two dashed horizontal lines shows the difference in the unadjusted means, which are also referenced by the centers of the marginal distributions for Memory.

> verdicts [that] assert that an individual's intelligence is a fixed quantity which cannot be increased. …. With practice, training, and above all method, we manage … to become more intelligent than we were before." (Binet, 1909/1975, pp. 106–107)

Intelligence as a performance measure was a reasonable position for Binet to espouse because his original test comprised items that poor learners had failed in school [IQ historians such as Deese (1993) have noted that Spearman's original tests included teacher ratings and grades in Latin; later IQ tests also included academic content but argued that IQ tests assess a person's learning capacity]. Burt (1937) presented the relation between IQ and achievement in terms of a container metaphor (Lakoff & Johnson, 1980), that of a jug.

> Capacity must obviously limit content. It is impossible for a pint jug to hold more than a pint of milk and it is equally impossible for a child's educational attainment to rise higher than his educable capacity. (Burt, 1937, p. 477)

This paradoxical view of aptitude assessment in which IQ is separate from learning outcome but independently measures learning potential has been termed "milk and jug" thinking (Share et al., 1989).

IQ Discrepancy Definitions of LD

In 1939, Thomson [cited in Deary et al. (2008)] pointed out that intelligence is not helpful in performing an academic test, even though it might have helped a candidate acquire the academic knowledge being tested. A less nuanced view, that LD is best defined by a concurrent discrepancy between IQ and achievement and should be defined in reference to IQ, was enshrined in U.S. special education regulations for LD from 1975 to 2004. Later, investigators concluded that IQ was largely irrelevant to the definition of LD (Siegel, 1992), and the U.S. special education regulations were modified in 2004 so that IQ tests could not be mandated for LD identification (Fletcher et al., 2007).

Analyses of the IQ–achievement discrepancy in LD show that IQ is not a proxy for learning potential. Francis et al. (1996) showed the weakness of the conceptual rationale for

https://doi.org/10.1017/S1355617709090481 Published online by Cambridge University Press

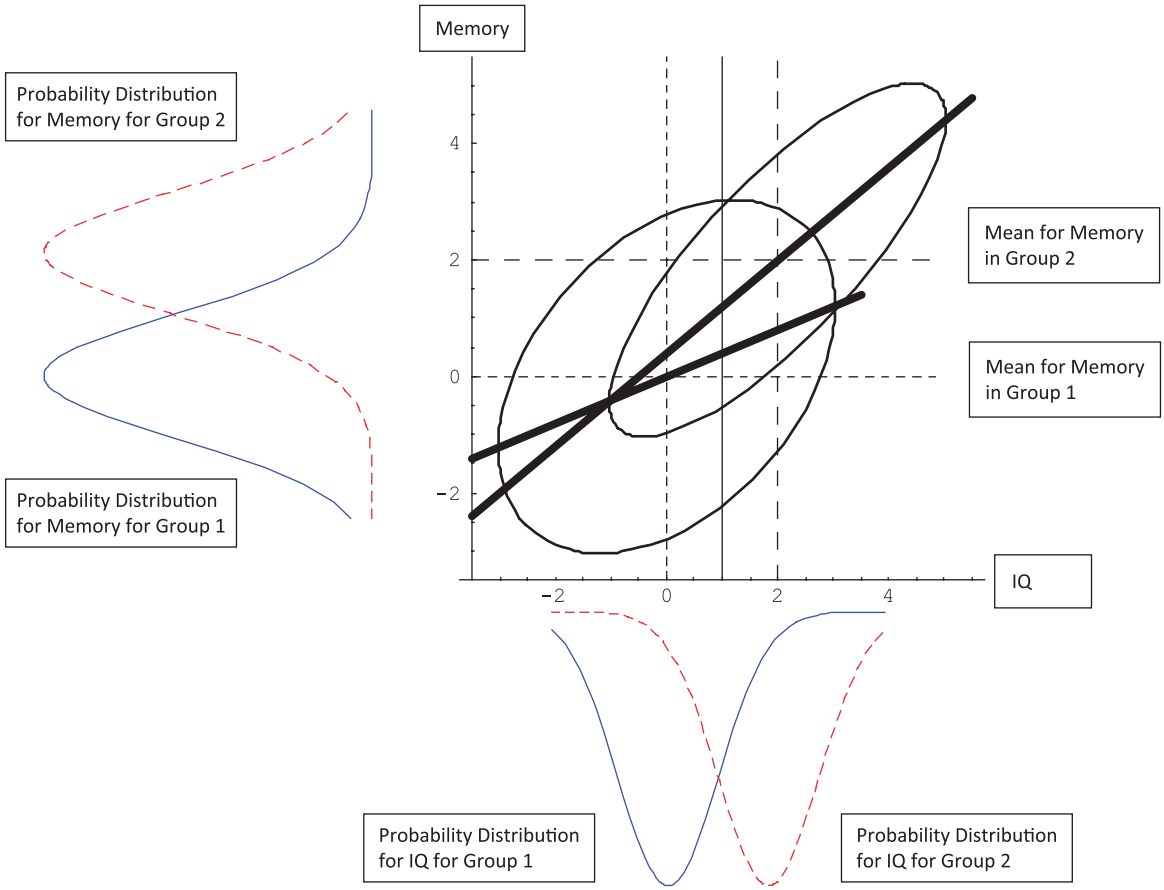

Fig. 4. In this figure, the ellipses are of different sizes reflecting the overall weaker relation between IQ and Memory in the lower performing group. In this case, the correlation between the two measures is only .4, whereas in the higher performing group, the correlation is .8.

models suggesting that IQ directly influences the attainment of academic and/or language skills, pointed out the limitations of the psychometric significance of IQ–attainment difference scores, and identified the limitations of simple comparisons of IQ and attainment measures. A meta-analytic study comparing cognitive functions in children with reading disabilities found only small effect size differences between poorer readers relative to discrepant and nondiscrepant IQ scores (Stuebing et al., 2002). IQ is a poor predictor of response to reading intervention (Fletcher et al., 2007; Mathes et al., 2005; Vellutino et al., 2000), and longitudinal studies have found no outcome differences between IQ-discrepant and IQ-nondiscrepant poor readers (Francis et al., 1996; Share et al., 1989).

IQ is itself influenced by many schooling differences (Ceci, 1991). Reduced word and print exposure in poor children or children who cannot read produces lowered IQ and learning over time, suggestive of what Stanovich (1986) termed a "Matthew effect" in which those who read well read more, and those who read poorly read less, leading to a long-term decline in reading and language skills. The influence of gene and environment is bidirectional in that the same developmental disadvantage that is part of many neurodevelopmental disorders lowers both IQ and academic skills (Hart & Risley, 1995).

IQ and ADHD

In 1908, Binet noted that the intelligence measured in his tests did not measure "the intelligence which is needed for … being attentive" (pp. 258–259). Later studies have confirmed the relatively weak association between ADHD and IQ (corresponding to 2–8 IQ points), and the mediation of any association by test-taking behavior, achievement deficits, and behavioral comorbidities (Bridgett & Walker, 2006; Fergusson et al., 1993; Frazier et al., 2004; Goodman et al., 1995; Jepsen et al., in press; Kuntsi et al., 2004; Rapport et al., 1999).

IQ and executive function are each associated with DRD4 and DAT1 risk alleles, both implicated in ADHD (Boonstra et al., 2008; Doyle et al., 2005; Khan & Faraone, 2006; Mill et al., 2006). However, group differences in executive function are not explained by group differences in IQ, or *vice versa*; IQ and executive function are not coheritable because correlations and sibling cross-correlations are not significant between executive function and IQ; deficits in each domain do not cosegregate within families; and there is independent familial segregation of both IQ and executive functions (Rommelse et al., 2008). Attempting to control for IQ differences when examining specific neuropsychological deficits like executive function in ADHD (Barkley

https://doi.org/10.1017/S1355617709090481 Published online by Cambridge University Press

et al., 2001; Murphy et al., 2001) is methodologically tenuous (Frazier et al., 2004) because decrements in overall ability are a feature of ADHD (and of any neurodevelopmental disorder defined in terms of cognitive-behavioral deficits), making statistical "control" impossible (Campbell & Kenny, 1999).

Controls: Matching for IQ and Selection Bias

The characteristics of controls will depend on the nature of the research question, populations, and exactly what the researchers want to control. While it is difficult to imagine a situation in which control of IQ was desirable if the comparison was with typically developing children, it may be desirable to control for sociodemographic characteristics that, in turn, are associated with higher than average IQ scores. When control IQ scores are elevated, a careful check should be made of sociodemographic characteristics in ascertainment bias.

Matching IQ to controls in children with a neurodevelopmental disorder (by child or by groups) creates unrepresentative groups. Either the neurodevelopmental disorder group will have higher IQs than the population with that disorder or the control group will have IQ scores below normative expectations. Comparison on a dependent variable that is correlated with IQ would lead to regression to the mean depending on which variable and group are compared (Campbell & Erlebacher, 1970).

Causal Inferences When IQ is a Covariate

Covarying for IQ may provide a comparison of groups (of typically developing children or children with a neurodevelopmental disorder) at values of the covariate that essentially do not exist in nature or are at best unrepresentative of the populations of interest, with a selection bias operating at the level of the sample (i.e., the process of sampling) or the population (i.e., the process by which members of the population are members of one subpopulation compared to another).

In the circumstances given above, including one where the mean IQ for the group with a neurodevelopmental disorder exceeds the mean IQ in the normative population, ANCOVA does not provide control for (or an interpretation of) the impact of IQ on other neurocognitive outcomes. Augmentation of group differences will occur when groups are compared on any measure that correlates positively with IQ, so covarying IQ cannot be used to "equate" the groups, which have been constructed in such a way as to make the groups nonequivalent in IQ.

Covariance analysis using IQ is usually predicated on the hypothesis that IQ "causes" the difference on a correlated variable (e.g., memory). When there is an inherent IQ difference between groups and the IQ difference is not separable from the level of the independent variable to which the patient belongs, the causal mechanism cannot be determined. The group difference in IQ remains a potential explanation for group differences on other cognitive measures and cannot be ruled out through statistical adjustment or explained away statistically, regardless of whether IQ is significant as a covariate or whether the differences on the dependent variables are significant. We suggest that covariance analysis does not permit causal statements about (or help sort out causal mechanisms of) IQ when the IQ difference is an inherent group characteristic.

The key issue is what represents an adequate explanation for the observed difference between groups on measure of cognitive performance; IQ is one possibility. Even if IQ accounted for all the variability in performance on a cognitive task, we cannot distinguish between IQ as a cause, IQ as an outcome, or a spurious association between IQ and the cognitive measures resulting from both tests measuring a common latent construct, in which case we would still have to identify the common latent variable and its relation to both IQ and the cognitive measure. When there is concern about the explanatory power of IQ, the researcher must be able to interpret the relation of IQ and the dependent variable, an effort supported by studies that seek to understand the construct validity of different dependent measures and their relations with IQ (e.g., Frazier et al., 2004).

After Adjusting for Barometric Pressure at Each Mountain's Highest Point, the Appalachians are Higher Than the Himalayas

IQ scores are positively correlated with family level of income, education, and other SES factors (Kaufman, 2001; Sattler, 1993). These relations complicate interpreting IQ when a preexisting IQ difference occurs in a disorder associated with lower SES.

Lead ingestion is associated with poverty and lower SES. Individuals cannot be randomly assigned to ingest lead and we cannot determine from ANCOVA whether lower IQ and/or lower SES is a result or a cause of lead exposure. Covarying for differences in SES variables, which is common in studies of lead effects, may lead to the paradoxical finding that lead has a nonlinear association with IQ, so that lower blood levels of lead are more strongly linked to IQ than are higher blood lead levels (Bowers & Beck, 2006). In simulation studies, covarying for education produced better performance in alcoholics than in controls (Adams et al., 1985); in a reading level match design, children with dyslexia had better orthographic processing skills than typically developing children (Siegel et al., 1995). In these examples, covarying for IQ or SES adjusted the means to levels not likely to be observed in nature or assumed a form of relationship between IQ and the outcome not supported by the data.

The brain systems with the most protracted postnatal development (e.g., the perisylvian areas important for language) are most susceptible to environmental influences and show the strongest associations with SES (Farah et al., 2006; Noble et al., 2005). It is important, therefore, to understand relations with environmental variables that are also associated with preexisting group differences.

https://doi.org/10.1017/S1355617709090481 Published online by Cambridge University Press

Jingles, Jangles, and Theoretically Vacuous Findings

In 1927, Kelley noted that theoretically meaningless findings arise from jingles (using the same term for different constructs) and jangles (using different terms for similar constructs). For example, IQ and achievement sound different because they have different "jangles," even though "the community between these two functions is nine times as great as the disparity between them" (Kelley, 1927, p. 63).

Under the hypothesis that IQ measures potential and capacity rather than performance and achievement, different jangles have been assigned to the same construct, depending on whether it formed part of an IQ or a cognitive battery. At one time, the Wechsler Intelligence Scales for Children (Wechsler, 1974, 1991) included repeating digits backward as part of how IQ was measured, while repeating digits backward in contemporaneous cognitive batteries was construed as working memory (e.g., Woodcock & Johnson, 1989). To covary an IQ measure that contained repeating digits backward from a task of repeating digits backward would constitute a jangle fallacy (Kelley, 1927).

Even when IQ correlates with an outcome variable, this relation is often theoretically vacuous because there is no specification of how IQ fits into a model of the cognitive function. In many neurocognitive studies, the theoretical model includes the dependent variable but not IQ. For example, children with SB have normal levels of single word decoding (hypothesized not to differ from age peers) but poor reading comprehension (hypothesized to differ from age peers), but there is no hypothesis about IQ, whether similar or different (Barnes & Dennis, 1992).

Many studies of neurodevelopmental disorders now include a theoretically relevant discriminant measure that differs from the dependent variable of interest in a specific, theory-relevant manner. Processes studied recently in SBM, for instance, have included saccadic adaptation (Salman et al., 2006), smooth pursuit eye movements (Salman et al., 2007), perception of timing intervals around a half-second (Dennis et al., 2004), inhibition of return (Dennis et al., 2005b), and mental model integration during language comprehension (Barnes et al., 2007). In studies of stimulus orienting in SB, the functional model involves intact top-down control *versus* impaired bottom-up control (Dennis et al., 2005a); the finding that groups with SBM can perform top-down but not bottom-up stimulus orienting is interpretable within the model, and without reference to IQ scores, whatever their levels. Processes studied recently in ADHD include post-error slowing (Schachar et al., 2004), response predictability (Aase et al., 2006), and cancellation and restraint inhibition (Schachar et al., 2007). Studies of response inhibition in ADHD measures response inhibition dynamically, adjusting the test parameters for each individual, so that the test measure—stop signal reaction time—is a within-individual measure, making IQ an inappropriate and/or irrelevant covariate for the specific cognitive functions of the ADHD cognitive phenotype.

In these examples, the appropriate control measures are the discriminant variables on which the neurodevelopmental groups do not differ from peers, and these, not IQ, facilitate a principled interpretation of group differences. To consider IQ scores as a covariate in these analyses would be to subtract the most general and theoretically impotent outcome measure from a tightly defined, highly specific, model-driven cognitive process. IQ has no place as a covariate in the statistical model for impaired performance because the IQ-adjusted model parameters are not the parameters of the theoretical model of performance. Having IQ "in the model" does not of itself afford a more precise answer to the question of whether differences in the construct of interest are caused by the neurodevelopmental process that differentiates the groups or whether group differences on the construct have theoretical importance for understanding the neurodevelopmental disorder.

IQ cannot be a discriminant measure in models of neurocognitive outcomes. To the extent that IQ represents the same processes as the construct of interest, then controlling for IQ removes variability in the outcome measure that is directly related to the construct of interest. Under such circumstances, IQ serves as a poor covariate, making any conclusions about specific cognitive processes more difficult and increasing interpretive complexity by removing some unspecified aspect of the dependent measure from itself. Even when the goal of including IQ as a covariate is to more clearly elucidate a theoretical question, frequently it either fails to do so, or it is less appropriate than alternative methods not including IQ at answering the question.

IQ in Childhood Acquired Conditions

We have discussed the use of IQ scores as covariates in neurodevelopmental disorders and concluded that it is generally inappropriate. The onset of many childhood disorders, however, occurs after a period of typical development. Children who develop strokes, brain tumors, leukemia, or who sustain anoxia, TBI, or other forms of childhood acquired brain insult, all have had preinsult time periods of variable length in which they developed normally.

There are two situations in which IQ might be considered to be a covariate in cases of childhood acquired brain insults. When preinjury IQ scores or IQ proxies are available, it is reasonable to use preinjury scores as covariates in considering the effects of postinjury measures of cognitive function. When IQ scores are derived postinjury, not preinjury, many of the same considerations apply that we have discussed for neurodevelopmental disorders. IQ scores obtained 1 year after a childhood TBI, for instance, will reflect the effects of the injury, and research suggests that the younger the child at the time of the injury, and the longer the time since the injury, the more cognitive measures will represent the effects of the injury. If either pre- or postinjury IQ scores is the proposed covariate, the requirement we outlined earlier, that any covariate, including IQ scores, should have a theoretically specified relation to the outcome measure, continues to hold.

https://doi.org/10.1017/S1355617709090481 Published online by Cambridge University Press

CONCLUSIONS

IQ scores have some value in the study of neurodevelopmental disorders. As products of multiple influences, they are useful, if volatile, indices of global functional outcome: the final common path of the child's genetic, biological, neural, cognitive, educational, and experiential life. IQ scores provide a general index of the representativeness of a sample, which facilitates comparisons of global outcomes across neurodevelopmental disorders.

Because IQ tests measure multiple, correlated abilities that often themselves correlate with dependent neuropsychological variables of interest, it is tempting to include IQ routinely in models of outcome. In the absence of an articulated model of function, IQ is a poorly specified latent variable that does not independently measure aptitude and potential or cause more specific cognitive processes, so that, generally, it should not be used as a covariate in investigating these processes.

IQ should be used as a covariate only in those rare circumstances where selection bias has produced problems of nonrepresentativeness in the sample, or where the theoretical model specifies its fit. If the group IQ is markedly deviant from expectations for the disorder, then some attempt to adjust for the sampling bias through IQ adjustment may be warranted to obtain a better estimate of the population mean of an outcome that is correlated with IQ. If the research question involves the link of IQ and a particular outcome, then approaches that involve construct validity, such as a latent variable approach, are likely to provide a better understanding of the phenomenon of interest.

As a field, neuropsychology needs more thoughtful use of IQ as a statistical adjustment in models of cognition. We hope that researchers and reviewers will consider the issues in this article before routinely recommending that IQ be controlled or covaried in studying neurodevelopmental disorders.

The idea that we require a theoretical model of cognition before understanding IQ is not new. An early statement of the idea, perhaps surprisingly, came from Spearman himself, the father of *g*: "No serviceable definition can possibly be found for general intelligence, until the entire psychology of cognition is established" (Spearman, 1923, p. 5). We concur.

ACKNOWLEDGMENTS

Preparation of this article was supported by National Institute of Child Health and Human Development Grants P01 HD35946 and P01 HD35946-06, "Spina Bifida: Cognitive and Neurobiological Variability." We thank Katia Sinopoli for helpful comments on the manuscript and Arianna Stefanatos for help with Manuscript preparation. We acknowledge an unidentified contributor to the pediatric neuropsychology list serve for the mountain comment.

REFERENCES

Aase, H., Meyer, A., & Sagvolden, T. (2006). Moment-to-moment dynamics of ADHD behaviour in South African children. *Behavioral and Brain Functions*, *2*, 11. doi: 10.1186/1744-9081-2-11.

Adams, K.M., Brown, G.G., & Grant, I. (1985). Analysis of covariance as a remedy for demographic mismatch of research subject groups: Some sobering simulations. *Journal of Clinical and Experimental Neuropsychology*, *7*, 445–462.

Barkley, R.A., Murphy, K.R., & Bush, T. (2001). Time perception and reproduction in young adults with attention-deficit/hyperactivity disorder. *Neuropsychology*, *15*, 351–360.

Barnes, M.A. & Dennis, M. (1992). Reading in children and adolescents after early onset hydrocephalus and in normally developing age peers: Phonological analysis, word recognition, word comprehension, and passage comprehension skill. *Journal of Pediatric Psychology*, *17*, 445–465.

Barnes, M.A., Huber, J., Johnston, A.M., & Dennis, M. (2007). A model of comprehension in spina bifida meningomyelocele: Meaning activation, integration, and revision. *Journal of the International Neuropsychological Society*, *13*, 854–864.

Binet, A. (1909/1975). *Modern ideas about children*; translated by S. Heisler. (Original work published 1909). Menlo Park, CA: Suzanne Heisler.

Binet, A. & Simon, T. (1916). *The development of intelligence in children (The Binet-Simon Scale)*; translated by E.S. Kite. Training School. Vineland, NJ: Suzanne Heisler.

Boonstra, A.M., Kooij, J.J.S., Buitelaar, J.K., Oosterlaan, J., Sergeant, J.A., Heister, J.G.A.M.A., & Franke, B. (2008). An exploratory study of the relationship between four candidate genes and neurocognitive performance in adult ADHD. *American Journal of Medical Genetics Part B (Neuropsychiatric Genetics)*, *147B*, 397–402.

Bowers, T.S. & Beck, B.D. (2006). What is the meaning of nonlinear dose-response relationships between blood lead concentrations and IQ? *Neurotoxicology*, *27*, 520–524.

Bridgett, D.J. & Walker, M.E. (2006). Intellectual functioning in adults with ADHD: A meta-analytic examination of full scale IQ differences between adults with and without ADHD. *Psychological Assessment*, *18*, 1–14.

Burt, C. (1937). *The backward child*. New York: Appleton-Century.

Campbell, D.T. & Erlebacher, A.E. (1970). How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In J. Hellmuth (Ed.), *Disadvantaged child: Vol. 3. Compensatory education: A national debate* (pp. 185–210). New York: Brunner/Mazel.

Campbell, D.T. & Kenny, D.A. (1999). *A primer on regression artifacts*. New York: Guilford Press.

Carroll, J.B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.

Cattell, R.B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, *40*, 153–193.

Ceci, S.J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology*, *27*, 703–722.

Craik, F.I.M. & Bialystok, E. (Eds.) (2006). *Lifespan cognition: Mechanisms of change*. New York: Oxford University Press.

Deary, I.J., Lawn, M., & Bartholomew, D.J. (2008). A conversation between Charles Spearman, Godfrey Thomson, and Edward L. Thorndike: The International Examinations Enquiry Meetings 1931-1938. *History of Psychology*, *11*, 122–142.

Deese, J. (1993). Human abilities versus intelligence. *Intelligence*, *17*, 107–116.

Dennis, M., Edelstein, K., Copeland, K., Frederick, J., Francis, D.J., Hetherington, R., Blaser, S.E. Kramer, L.A., Drake, J.M., Brandt, M., & Fletcher, J.M. (2005a). Covert orienting to exogenous

https://doi.org/10.1017/S1355617709090481 Published online by Cambridge University Press

and endogenous cues in children with spina bifida. *Neuropsychologia*, *43*, 976–987.

Dennis, M., Edelstein, K., Copeland, K., Frederick, J.A., Francis, D.J., Hetherington, R., Blaser, S.E., Kramer, L.A., Drake, J.M., Brandt, M.E., & Fletcher, J.M. (2005b). Space-based inhibition of return in children with spina bifida. *Neuropsychology*, *19*, 456–465.

Dennis, M., Edelstein, K., Hetherington, R., Copeland, K., Frederick, J., Blaser, S.E., Kramer, L.A., Drake, J.M., Brandt, M., & Fletcher, J.M. (2004). Neurobiology of perceptual and motor timing in children with spina bifida in relation to cerebellar volume. *Brain*, *127*, 1293–1301.

Doyle, A.E., Willcutt, E.G., Seidman, L.J., Biederman, J., Chouinard, V.-A., Silva, J., & Faraone, S.V. (2005). Attention-deficit/hyperactivity disorder endophenotypes. *Biological Psychiatry*, *57*, 1324–1335.

Evans, B. & Waites, B. (1981). *IQ and mental testing: An unnatural science and its social history*. London: Macmillan Press.

Evans, S.H. & Anastasio, E.J. (1968). Misuse of analysis of covariance when treatment effect and covariate are confounded. *Psychological Bulletin*, *69*, 225–234.

Farah, M.J., Shera, D.M., Savage, J.H., Betancourt, L., Giannetta, J.M., Brodsky, N.L., Malmud, E.K., & Hurt, H. (2006). Childhood poverty: Specific associations with neurocognitive development. *Brain Research*, *1110*, 166–174.

Fergusson, D.M., Horwood, L.J., & Lynskey, M.T. (1993). The effects of conduct disorder and attention deficit in middle childhood on offending and scholastic ability at age 13. *Journal of Child Psychology and Psychiatry*, *34*, 899–916.

Fletcher, J.M., Lyon, G.R., Fuchs, L., & Barnes, M. (2007). *Learning disabilities: From identification to intervention*. New York: Guilford Press.

Flynn, J.R. (2007). *What is intelligence? Beyond the Flynn Effect*. New York: Cambridge University Press.

Francis, D.J., Fletcher, J.M., Shaywitz, B.A., Shaywitz, S.E., & Rourke, B.P. (1996). Defining learning and language disabilities: Conceptual and psychometric issues with the use of IQ tests. *Language, Speech, and Hearing Services in Schools*, *27*, 132–143.

Frazier, T.W., Demaree, H.A., & Youngstrom, E.A. (2004). Meta-analysis of intellectual and neuropsychological test performance in attention-deficit/hyperactivity disorder. *Neuropsychology*, *18*, 543–555.

Goddard, H.H. (1917). Mental tests and immigrants. *Journal of Delinquency*, *2*, 243–277.

Goodenough, F.L. (1949). *Mental testing*. New York: Rinehart.

Goodman, R., Simonoff, E., & Stevenson, J. (1995). The impact of child IQ, parent IQ and sibling IQ on child behavioural deviance scores. *Journal of Child Psychology and Psychiatry*, *36*, 409–425.

Gottfredson, L.S. (2004). Intelligence: Is it the epidemiologists' elusive "fundamental cause" of social class inequalities in health? *Journal of Personality and Social Psychology*, *86*, 174–199.

Gould, S.J. (1981). *The mismeasure of man*. New York: Norton.

Haier, R.J., Jung, R.E., Yeo, R.A., Head, K., & Alkire, M.T. (2004). Structural brain variation and general intelligence. *NeuroImage*, *23*, 425–433.

Haier, R.J., Jung, R.E., Yeo, R.A., Head, K., & Alkire, M.T. (2005). The neuroanatomy of general intelligence: Sex matters. *NeuroImage*, *25*, 320–327.

Hart, B. & Risley, T.R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Brookes.

Hebb, D.O. (1949). *The organization of behavior*. New York: Wiley.

Hegarty, P. (2007). From genius inverts to gendered intelligence: Lewis Terman and the power of the norm. *History of Psychology*, *10*, 132–155.

Horn, J.L. (1998). A basis for research on age differences in cognitive capabilities. In J.J. McArdle & R.W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 57–91). Mahwah, NJ: Lawrence Erlbaum.

Huitema, B.E. (1980). *The analysis of covariance and alternatives*. New York: Wiley.

Jensen, A.R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, *39*, 1–123.

Jensen, A.R. (1985). The nature of the black-white difference on various psychometric tests: Spearman's hypothesis. *The Behavioral and Brain Sciences*, *8*, 193–263.

Jensen, A.R. (1989). Raising IQ without raising *g*? A review of the Milwaukee project: Preventing mental retardation in children at risk. *Developmental Review*, *9*, 234–258.

Jensen, A.R. & Weng, L.-J. (1994). What is a good *g? Intelligence*, *18*, 231–258.

Jepsen, J.R.M., Fagerlund, B., & Mortensen, E.L. (in press). Do attention deficits influence IQ assessment in children and adolescents with ADHD? *Journal of Attention Disorders*. Doi:10.1177/1087054708322996.

Johnson, W. & Bouchard, T.J. (2005). The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence*, *33*, 393–416.

Johnson, W., Jung, R.E., Colom, R., & Haier, R.J. (2008). Cognitive abilities independent of IQ correlate with regional brain structure. *Intelligence*, *36*, 18–28.

Jung, R.E. & Haier, R.J. (2007). The parieto-frontal integration theory (P-FIT) of intelligence: Converging neuroimaging evidence. *Behavioral and Brain Sciences*, *30*, 135–154.

Kamin, L.J. (1974). *The science and politics of IQ*. Potomac, MD: Lawrence Erlbaum Associates.

Kane, H. & Oakland, T.D. (2000). Secular declines in Spearman's g: Some evidence from the United States. *Journal of Genetic Psychology*, *161*, 337–345.

Kaufman, A.S. (2001). Do low levels of lead produce IQ loss in children? A careful examination of the literature. *Archives of Clinical Neuropsychology*, *16*, 303–341.

Kelley, T.L. (1927). *Interpretation of educational measurements*. Yonkers-on-Hudson, New York: World Book Company.

Khan, S.A. & Faraone, S.V. (2006). The genetics of ADHD: A literature review of 2005. *Current Psychiatry Reports*, *8*, 393–397.

Kornhaber, M., Krechevsky, M., & Gardner, H. (1990). Engaging intelligence. *Educational Psychologist*, *25*, 177–199.

Kuntsi, J., Eley, T.C., Taylor, A., Hughes, C., Asherson, P., Caspi, A., & Moffitt, T.E. (2004). Co-occurrence of ADHD and low IQ has genetic origins. *American Journal of Medical Genetics Part B (Neuropsychiatric Genetics)*, *124B*, 41–47.

Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Lord, F.M. (1965). A strong true-score theory, with applications. *Psychometrika, 30*, 239–270.

Lord, F.M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin*, *68*, 304–305.

Lord, F.M. (1969). Statistical adjustments when comparing preexisting groups. *Psychological Bulletin*, *72*, 336–337.

Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) "'General intelligence,' objectively determined and measured." *Journal of Personality and Social Psychology*, *86*, 96–111.

https://doi.org/10.1017/S1355617709090481 Published online by Cambridge University Press

Lynn, R. (1978). Ethnic and racial differences in intelligence: International comparisons. In R.T. Osbourne, C.E. Noble & N. Weyl (Eds.), *Human variation: The biopsychology of age, race, and sex*, (pp. 261–286). New York: Academic Press.

Mathes, P.G., Denton, C.A., Fletcher, J.M., Anthony, J.L., Francis, D.J., & Schatschneider, C. (2005). The effects of theoretically different instruction and student characteristics on the skills of struggling readers. *Reading Research Quarterly*, *40*, 148–182.

Mill, J., Caspi, A., Williams, B.S., Craig, I., Taylor, A., Polo-Tomas, M., Berridge, C.W., Poulton, R., & Moffitt, T.E. (2006). Prediction of heterogeneity in intelligence and adult prognosis by genetic polymorphisms in the dopamine system among children with attention-deficit/hyperactivity disorder: Evidence from 2 birth cohorts. *Archives of General Psychiatry*, *63*, 462–469.

Miller, G.A. & Chapman, J.P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology*, *110*, 40–48.

Murphy, K.R., Barkley, R.A., & Bush, T. (2001). Executive functioning and olfactory identification in young adults with attention deficit-hyperactivity disorder. *Neuropsychology*, *15*, 211–220.

Neisser, U., Boodoo, G., Bouchard, T.J, Boykin, A.W., Brody, N., Ceci, S.J., Halpern, D.F., Loehlin, J.C., Perloff, R., Sternberg, R.J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*, 77–101.

Noble, K.G., Norman, M.F., & Farah, M.J. (2005). Neurocognitive correlates of socioeconomic status in kindergarten children. *Developmental Science*, *8*, 74–87.

Piéron, H. (1932). Theoretical and practical aspects of intelligence. *British Journal of Psychology*, *22*, 353–358.

Porteus, S.D. (1917). Mental tests with delinquents and aboriginal children. *Psychological Review*, *24*, 32–41.

Rapport, M.D., Scanlan, S.W., & Denney, C.B. (1999). Attention-deficit/hyperactivity disorder and scholastic achievement: A model of dual developmental pathways. *Journal of Child Psychology and Psychiatry*, *40*, 1169–1183.

Roll-Hansen, N. (1988). The progress of eugenics: Growth of knowledge and change in ideology. *History of Science*, *26*, 295–331.

Rommelse, N.N., Altink, M.E., Oosterlaan, J., Buschgens, C.J., Buitelaar, J., & Sergeant, J.A. (2008). Support for an independent familial segregation of executive and intelligence endophenotypes in ADHD families. *Psychological Medicine*, *8*, 1595–1606.

Salman, M.S., Sharpe, J.A., Eizenman, M., Lillakas, L., To, T., Westall, C., Steinbach, M.J., & Dennis, M. (2006). Saccadic adaptation in Chiari Type II malformation. *Canadian Journal of Neurological Sciences*, *33*, 372–378.

Salman, M.S., Sharpe, J.A., Lillakas, L., Steinbach, M.J., & Dennis, M. (2007). Smooth ocular pursuit in Chiari type II malformation. *Developmental Medicine and Child Neurology*, *49*, 289–293.

Sattler, J.M. (1993). *Assessment of children's intelligence and special abilities*. New York: Allyn & Bacon.

Schachar, R.J., Chen, S., Logan, G.D., Ornstein, T.J., Crosbie, J., Ickowicz, A., & Pakulak, A. (2004). Evidence for an error monitoring deficit in attention deficit hyperactivity disorder. *Journal of Abnormal Child Psychology*, *32*, 285–293.

Schachar, R., Logan, G.D., Robaey, P., Chen, S., Ickowicz, A., & Barr, C. (2007). Restraint and cancellation: Multiple inhibition deficits in attention deficit hyperactivity disorder. *Journal of Abnormal Child Psychology*, *35*, 229–238.

Schneider, W.H. (1992). After Binet: French intelligence testing, 1900-1950. *Journal of the History of the Behavioural Sciences*, *28*, 111–132.

Share, D.L., McGee, P.A., & Silva, P.A. (1989). IQ and reading progress: A test of the capacity notion of IQ. *Journal of the American Academy of Child and Adolescent Psychiatry*, *28*, 97–100.

Siegel, L.S. (1992). An evaluation of the discrepancy definition of dyslexia. *Journal of Learning Disabilities*, *25*, 618–629.

Siegel, L.S., Share, D., & Geva, E. (1995). Evidence for superior orthographic skills in dyslexics. *Psychological Science*, *6*, 250–254.

Siegler, R.S. (1992). The other Alfred Binet. *Developmental Psychology*, *28*, 179–190.

Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology*, *15*, 201–293.

Spearman, C. (1914). Theory of two factors. *Psychological Review*, *21*, 105–115.

Spearman, C. (1923). *The nature of "intelligence" and the principles of cognition*. London: Macmillan.

Spearman, C. (1927). *The ability of man: Their nature and measurement*. New York: Macmillan.

Stanovich, K.E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*, 360–407.

Stuebing, K.K., Fletcher, J.M., LeDoux, J.M., Lyon, G.R., Shaywitz, S.E., & Shaywitz, B.A. (2002). Validity of IQ-discrepancy classifications of reading disabilities: A meta-analysis. *American Educational Research Journal*, *39*, 469–518.

Teasdale, T.W. & Owen, D.R. (2008). Secular declines in cognitive test scores: A reversal of the Flynn effect. *Intelligence*, *36*, 121–126.

Terman, L.C. (1916). *The Stanford revision of the Binet-Simon test*. Boston, MA: Houghton Mifflin.

Thomson, G.H. (1916). A hierarchy without a general factor. *British Journal of Psychology*, *8*, 271–281.

Thomson, G.H. (1919). The hierarchy of abilities. *British Journal of Psychology*, *9*, 337–344.

Tupper, D.E. & Rosenblood, L.K. (1984). Methodological considerations in the use of attribute variables in neuropsychological research. *Journal of Clinical and Experimental Neuropsychology*, *6*, 441–453.

Vellutino, F.R., Scanlon, D.M., & Lyon, G.R. (2000). Differentiating between difficult-to-remediate and readily remediated poor readers. *Journal of Learning Disabilities*, *33*, 223–238.

Vernon, P.E. (1964). *Personality assessment: A critical survey*. London, Methuen.

Wechsler, D. (1974). *Wechsler Intelligence Scale for Children-Revised*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children—Third Edition*. San Antonio, TX: Psychological Corporation.

Wilke, M., Sohn, J.-H., Byars, A.W., & Holland, S.K. (2003). Bright spots: Correlations of gray matter volume with IQ in a normal pediatric population. *NeuroImage*, *20*, 202–215.

Wolf, T.H. (1973). *Alfred Binet*. Chicago, IL: The University of Chicago Press.

Woodcock, R.W. & Johnson, M.B. (1989). *Tests of cognitive ability*. New York: Riverside Publishing.

https://doi.org/10.1017/S1355617709090481 Published online by Cambridge University Press