Article:
von Stumm, Sophie orcid.org/0000-0002-0447-5471 and Nancarrow, Allie F. (2024) New methods, persistent issues, and one solution:Gene-environment interaction studies of childhood cognitive development. Intelligence. 101834. ISSN: 0160-2896

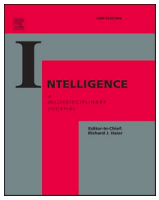
https://doi.org/10.1016/j.intell.2024.101834

Intelligence 105 (2024) 101834

Contents lists available at ScienceDirect

Intelligence

journal homepage: www.elsevier.com/locate/intell



New methods, persistent issues, and one solution: Gene-environment interaction studies of childhood cognitive development

Sophie von Stumm [a,*], Allie F. Nancarrow [b]

[a] Department of Education, University of York, United Kingdom
[b] Institute for Employment Studies, United Kingdom

ARTICLE INFO

Keywords:
Gene-environment interaction
Statistical power
Sample size
Effect size
Cognitive development

ABSTRACT

Children's differences in cognitive development stem from the complex interplay of genetic and environmental factors. Identifying gene-environment interactions in cognitive development is key for effectively targeting interventions that improve children's life chances. The advent of polygenic scores, which aggregate DNA variants to index a person's genetic propensities for phenotypic development, has created unprecedented opportunities for pinpointing gene-environment interactions. Yet, the issue of statistical power – the probability of detecting a true effect – prevails, and no replicable gene-environment interactions in child cognitive development have been reported. In this review article, we recapitulate three approaches to studying gene-environment interactions, including twin studies, candidate gene models, and polygenic score methods. We then discuss the issue of statistical power in gene-environment interaction research and conclude that larger samples are key to ushering a new era of replicable gene-environment interaction findings.

Children's differences in cognitive development – the ability to reason, think, and learn – are evident early on in life (e.g., von Stumm, Kandaswamy, & Maxwell, 2023), and they have pervasive, long-term influence on many important life outcomes, including educational attainment, career success, health, and longevity (Deary, 2012). Understanding the causes of children's differences in cognitive development is therefore key to designing and implementing effective interventions that improve children's life chances (Protzko, Aronson, & Blair, 2013).

Children's differences in cognitive development stem from the complex interplay of genetic and environmental factors. Although correlations of genes and environments are widespread (Krapohl et al., 2017), no replicable gene-environment interactions have been demonstrated in the prediction of children's cognitive development (cf., Tucker-Drob & Bates, 2016; von Stumm et al., 2023). A central problem in detecting gene-environment interactions is statistical power, because interaction terms tend to have very small effect sizes, which require, in turn, very large sample sizes to be observed. For example, a 30-year review of the size of moderating effects of categorical variables in a regression model framework found a median observed effect size (f^2) of 0.002 (Aguinis, Beaty, Boik, & Pierce, 2005), which is a hundred times smaller than the effect size typically considered 'small' (i.e., 0.2; Cohen, 1988). Although power calculations are widely recognized as an important step in designing rigorous research studies (van der Sluis, Dolan, Neale, & Posthuma, 2008), hard-and-fast recommendations for statistical power when testing gene-environment interactions are scarce. To address this gap, we recapitulate here approaches to studying gene-environment interactions, including twin studies, candidate gene models, and polygenic score methods. We then review the issue of statistical power in gene-environment interaction research and conclude that there is one panacea – increasing sample sizes. We close with recommendations for growing sample sizes in interaction studies in developmental science.

1. Gene-environment interactions in cognitive development

A developmental gene by environment interaction describes when the effect of the genotype on phenotypic development depends on the environment, and/or when the effect of the environment on phenotypic development depends on the genotype (von Stumm et al., 2023). The direction of an interaction is statistically ambiguous: A gene-environment interaction model cannot determine per se whether the influence of genetics varies as a function of the environment, or whether the influence of the environment varies as a function of genetics.

* Corresponding author.
E-mail address: sophie.vonstumm@york.ac.uk (S. von Stumm).

https://doi.org/10.1016/j.intell.2024.101834
Received 26 October 2023; Received in revised form 28 March 2024; Accepted 24 April 2024
Available online 9 May 2024

Gene-environment interactions are typically interpreted according to one of three conceptual frameworks (von Stumm et al., 2023). In the first framework, called diathesis-stress model (Meehl, 1962; Sigelman & Rider, 2009), individuals with a higher genetic risk for poor developmental outcomes may be disproportionately negatively affected by environmental adversity, such as living in chaotic family homes or experiencing insufficient cognitive stimulation (Fig. 1a). Here, the negative effect of an environmental stressor on phenotypic development is weaker in individuals with lower genetic vulnerability, but stronger in individuals with greater genetic vulnerability (Zuckerman, 1999). For example, children with weaker genetic propensities for verbal ability may suffer significant developmental disadvantages if they are raised without many opportunities for joint book reading (cf., Noble et al., 2019), while children with stronger genetic propensities would not be much affected by a lack of joint book reading experiences.

Second, enriched environments may maximise the expression of genetic differences, while scant environments mask them (Fig. 1b). To come back to the example of verbal ability and book reading, the effects of a higher genetic propensity toward verbal ability for language development may be augmented in environments with more joint book reading opportunities but diminished when environmental stimulation in terms of joint booking reading is lacking (cf., bioecological model of human development; Bronfenbrenner & Morris, 2006; Pluess & Belsky, 2013).

The third framework suggests that individuals with greater genetic susceptibility may be disproportionately affected by both positive and negative environments (Greven et al., 2019). For example, more frequent joint book reading may greatly benefit language development in children who inherited a greater sensitivity to the environment, while less book reading may badly impair language development in the same children (Fig. 1c). This notion is central to theories of environmental sensitivity, which propose that individuals differ in their sensitivity to both aversive as well as supportive environments (Belsky & Pluess, 2009).

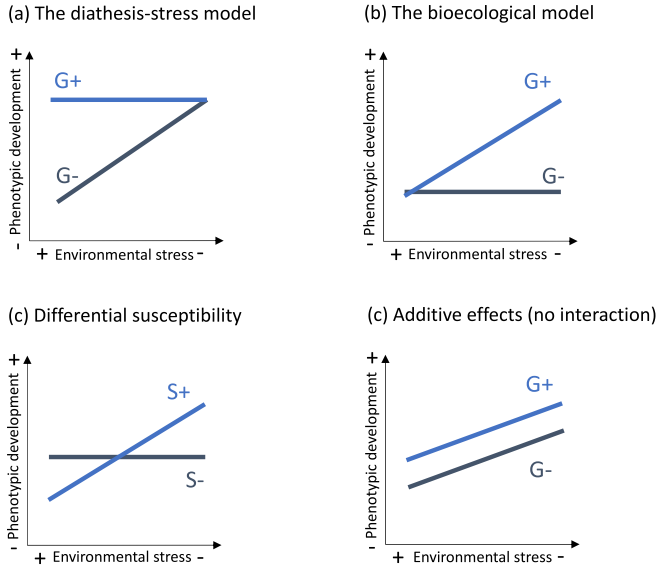


Fig. 1. Models of the gene-environment interplay.
Note. Panel (a) illustrates the diathesis-stress model, where high environmental stress (signified by +) has detrimental effects on the phenotypic development of individuals with greater genetic risk (G-) but not those with low genetic vulnerability (G+). Panel (b) shows the bioecological model, where environmental stress masks genetic differences, so that genetic differences become phenotypically expressed in positive (i.e., low environmental stress, signified by -) but not in negative environments. Panel (c) demonstrates differential susceptibility, where high and low stress environments have disproportionately larger effects on individuals with greater sensitivity to environmental influences (S+) compared to those with low environmental sensitivity (S-). Panel (d) shows additive effects of genetic and environmental factors without gene-environment interaction.

Gene-environment interactions are key to explaining why children differ in their responses to interventions that seek to improve their cognitive development (Asbury, McBride, & Rimfeld, 2021; Sokolowski & Ansari, 2018). Specifically, finding gene-environment interactions in cognitive development will help to identify the children who will benefit most from interventions that boost cognitive growth (Asbury et al., 2021). Conversely, finding no gene-environment interactions in cognitive development would indicate that intervention effects do not vary systematically as a function of children's genotypes (Fig. 1d). In this case, interventions that seek to improve cognitive development could be expected to be broadly effective for all children independent of their genetic differences.

2. Gene-environment interactions in twin studies

Twin and family studies estimate the extent to which individual differences in phenotypic traits can be attributed to genetic and environmental influences by comparing the phenotypic resemblance of individuals of differing degrees of genetic relatedness. For example, finding that monozygotic twins, who are genetically identical, show greater resemblance in cognitive development within pairs than dizygotic twins, who only share on average 50% of their segregating genes, suggests that differences in cognitive development are at least partly heritable.

Twin studies test for gene-environment interactions by estimating whether the heritability of a trait varies as a function of an environmental measure. A famous gene-environment interaction in childhood cognitive development that has been tested in twin studies is known as the Scarr-Rowe hypothesis. This hypothesis proposes that the heritability of children's intelligence is reduced in families with fewer socioeconomic resources compared to children raised in families with greater resources (Rowe, Jacobson, & Van den Oord, 1999; Scarr-Salapatek, 1971). The first empirical test of this hypothesis (Scarr-Salapatek, 1971) was criticised for its data and analytical approaches (Eaves & Jinks, 1972). Among other points, the critique outlined that up to 3500 twin pairs were needed to achieve adequate statistical power for testing the Scarr-Rowe hypothesis (Eaves & Jinks, 1972). Few of the subsequent studies that sought to demonstrate the Scarr-Rowe hypothesis met or exceeded these sample size demands (Tucker-Drob & Bates, 2016), including the one that contributed to the hypothesis' name (Rowe et al., 1999). Yet, one such study attracted much public and scientific interest (Turkheimer, Haley, Waldron, d'Onofrio, & Gottesman, 2003), accumulating over 1900 citations to date and triggering numerous replication attempts, many of which were not successful (Figlio, Freese, Karbownik, & Roth, 2017; Hanscombe et al., 2012; see Tucker-Drob and Bates (2016) for a meta-analysis). This seminal study analysed data from 114 monozygotic and 205 dizygotic American twin pairs, who completed an intelligence test at age 7 years, and whose family socioeconomic status (SES) was inferred from their parents' education, occupation, and income (Turkheimer et al., 2003). In children from higher SES families, the heritability of intelligence was around 60%, but in children from families of lower SES, the heritability of intelligence was close to zero (Turkheimer et al., 2003). These results were interpreted to mean that children in poverty do not get to develop their full genetic potential and thus, that environmental interventions could have large effects on these children's cognitive development (p. 134, Nisbett et al., 2012).

Samples of at least 3300 to 5000 twin pairs are needed to achieve sufficient statistical power for robustly detecting a gene-environment interaction of medium effect size, as reported by Turkheimer et al. (2003); ~0.20–0.30; Eaves & Jinks, 1972; Hanscombe et al., 2012; Tucker-Drob & Bates, 2016). Yet, twin samples of this magnitude are rarely available: A comprehensive meta-analysis of 14 studies on the

Scarr-Rowe hypothesis included only four independent samples that approximated the required number of twin pairs (Tucker-Drob & Bates, 2016), including two from the US with 2494 and 3203 twin pairs, respectively (Grant et al., 2010; Kirkpatrick, McGue, & Iacono, 2015), one from the UK with 8716 twin pairs (Hanscombe et al., 2012), and one from then Netherlands with 3132 twin pairs (Bartels, van Beijsterveldt, & Boomsma, 2009). None of these studies found conclusive evidence to support the Scarr-Rowe hypothesis (Bartels et al., 2009; Grant et al., 2010; Hanscombe et al., 2012; Kirkpatrick et al., 2015). Likewise, neither of the two studies with the largest number of twin pairs that have tested the Scarr-Rowe hypothesis to date (~ 9000 and ~ 25,000 twin pairs from the UK and US, respectively) reported evidence for a significant moderation of the heritability of intelligence by family background (Figlio et al., 2017; Hanscombe et al., 2012). Notwithstanding, the aforementioned meta-analysis concluded that the Scarr-Rowe hypothesis held up in samples from the US but neither in other countries (i.e., Australia, England and Wales, Germany, The Netherlands, Sweden) nor in the overall meta-analytic model (Tucker-Drob & Bates, 2016). While the meta-analytic findings in US samples were robust against a series of checks, the pooled estimates continued to show significant heterogeneity even after accounting for the effect of the sample's country, as well as for other moderators (e.g., age of assessment; Tucker-Drob & Bates, 2016). This heterogeneity reflects unexplained inconsistencies in results across studies that questions the validity of the pooled estimates.

3. Gene-environment interactions in candidate-gene approaches

While twin studies are pivotal for disentangling genetic and environmental effects, they cannot determine which genetic or environmental factors bring about phenotypic differences. Instead, twin studies attribute proportions of population variance to genetic versus environmental factors based on comparisons of the phenotypic resemblance of identical and fraternal twins. The candidate-gene approach to gene-environment interactions tests whether associations between variations (i.e., alleles) of a-priori-selected (i.e., candidate) genes and a given phenotype vary under different environmental conditions (Caspi et al., 2003; Dick et al., 2015). Many of these studies tested interactions between variants (e.g., 5-HTT, MAOA, DRD2, and COMT) and environmental stressors in the prediction of psychiatric outcomes, such as depression and schizophrenia (Dick et al., 2015; Duncan & Keller, 2011). The candidate-gene approach was based on the (false) premise that selected genetic variants or loci were strongly associated with phenotypic differences in complex traits, like cognitive development or psychiatric disorders. Today, there is broad consensus that developmental phenotypes show polygenicity, meaning they are influenced by a vast myriad of genetic variants, each of which exerts a minuscule effect size (i.e., Chabris, Lee, Cesarini, Benjamin, & Laibson, 2015; Okbay et al., 2022). For example, at least 3952 genetic variants are significantly associated with people's differences in the number of years that they spend in full-time education, and each of these variants explains a tiny proportion of variance in the target phenotype (Okbay et al., 2022).

Gene-environment interaction studies that relied on the candidate-gene approach were in retrospect underpowered (Chabris et al., 2015; Okbay et al., 2022). Their average sample size produced statistical power of about 10% (Duncan & Keller, 2011), which means that only 1 in 10 studies could detect a true interaction effect (Cohen, 1992), and that most significant candidate gene interaction effects are likely false positives (Chabris et al., 2015; Dick et al., 2015).

4. Gene-environment interactions in polygenic score approaches

The advent of polygenic scores, which aggregate inherited DNA variants to index a person's genetic propensity for a target phenotype, offered new opportunities for testing the interplay between genes and environments. Polygenic scores, also known as polygenic risk scores or polygenic indexes, are aggregates of common single-nucleotide polymorphisms (SNPs), whose associations with a target phenotype were identified in a genome-wide association (GWA) study (e.g., Okbay et al., 2022). Based on the summary statistics from GWA studies, polygenic scores can be computed for participants in independent samples, for whom genotype data are available, to predict phenotypic outcomes (Plomin & von Stumm, 2018). For example, polygenic scores based on the summary statistics from the most recent GWA study for years spent in full-time education predicted 14.3% of variance in this phenotype across two independent samples (Okbay et al., 2022). The prediction of polygenic scores is typically less than the heritability estimates reported in twin studies, in part because polygenic scores include only common SNPs (i.e., SNPs that occur in at least 1% of the population), in part because current GWA sample sizes are too small to identify all SNPs that are associated with a target phenotype, even when they are common, and in part because heritability estimates from twin studies include unmodelled gene-environment correlations. Yet, polygenic scores continue to gain in predictive validity, because GWA sample sizes are becoming larger, and because of methodological advances that optimise leveraging GWA summary statistics (e.g., Demange et al., 2021; Grotzinger et al., 2019).

It is possible to test for gene-environment interactions in multiple regression models by computing the product term of an environmental factor and the polygenic scores for a target phenotype, akin to any other interaction terms from observed measures. Such models allow simultaneously testing main effects of multiple environments and their interactions with polygenic scores, while controlling for gene-environment correlations (e.g., von Stumm et al., 2023). The term "polyenvironicity" has been proposed to describe the finding that numerous environmental factors are associated with differences in phenotypic development rather than one or two environments that excert large effect sizes (Cattan et al., 2022; von Stumm & d'Apice, 2022). Thus, polyenvironicity is analogous to polygenecity: Genetic and environmental influences on phenotypic development both result from numerous variants – may it be SNPs or environmental factors – that each have small independent effects (Götz, Gosling, & Rentfrow, 2022; von Stumm & d'Apice, 2022). The issue of polyenvironicity becomes even more pertinent when we consider that few environments are truly exogenous to the individual (i.e., independent of genetics), a problem known as endogeneity (Boardman, Daw, & Freese, 2013), which complicates disentangling environmental from genetic effects. Because many environmental measures show substantial genetic influences, it is plausible that environments and a target phenotype – here cognitive development – have shared genetic aetiology.

5. Statistical power

Statistical power is the probability of correctly rejecting the null hypothesis (Cohen, 1992). By convention, a study is considered adequately powered if the power to correctly reject the null hypothesis is 80% or above, meaning the null hypothesis will be correctly rejected in 4 out of 5 statistical tests. Low statistical power impairs the accumulation of scientific evidence because it increases the rate of false negatives and false positives and produces inflated effect sizes (Button et al., 2013).

Three factors are key to estimating statistical power. The first is the alpha level, which indexes the maximum *p*-value for which the null hypothesis will be rejected, that is, the maximum rate of false positives, which is conventionally set at $\alpha = 0.05$ or $\alpha = 0.01$ (Cohen, 1992). Raising the alpha level widens the range of *p*-values for which the null hypothesis will be rejected and thus, increases statistical power (Cohen, 1992). Because increasing the alpha level also augments the probability of false positives (Cohen, 1992), elevating the alpha level for the purpose of strengthening statistical inference is counterproductive.

The second factor is effect size, which describes the difference between the null and the alternative hypothesis (Cohen, 1992). The larger the effect size, the smaller the sample that is required to achieve the

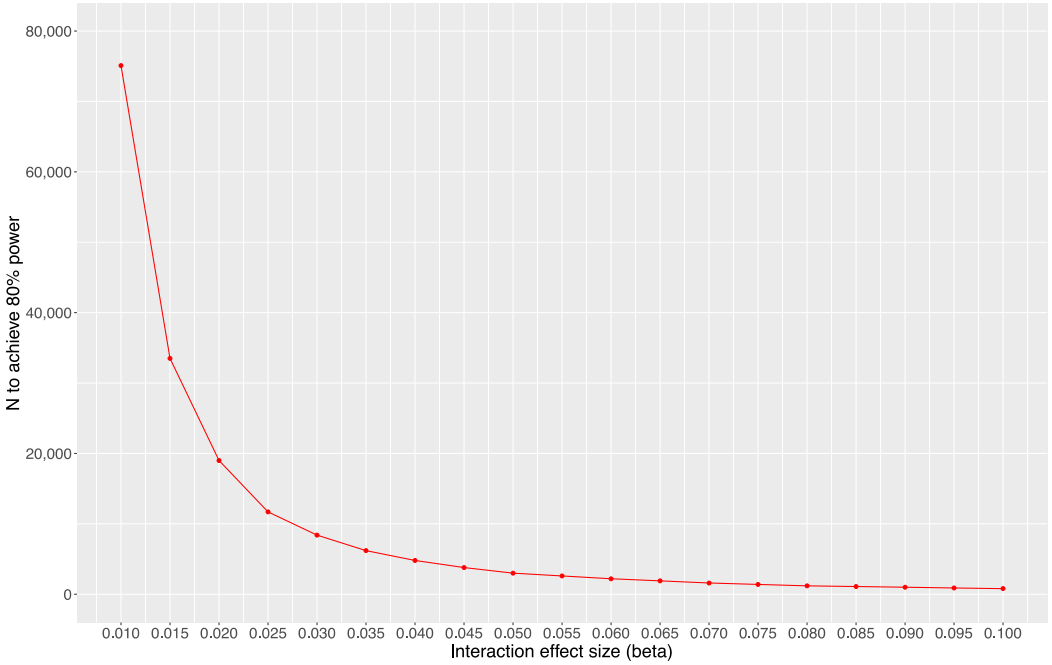


Fig. 2. Sample sizes to achieve 80% statistical power for identifying interaction effect sizes from $\beta = 0.01$ to 0.10.
Note. Power analyses were performed using InteractionPoweR (Baranger, 2019; Baranger et al., 2021). Direct effects, and the correlation between predictors, were specified at $r = 0.1$. All analyses including 1000 simulations and required sample sizes were rounded up to the nearest 100 participants.

statistical power for correctly rejecting the null hypothesis, because the easier it is to observe the effect (Abraham & Russell, 2008; Cohen, 1992). The third factor is therefore sample size. Larger samples tend to be more representative of studies' target populations and thus, achieve greater statistical power to detect even small effect sizes with low standard errors (Cohen, 1992).

The issue of statistical power applies to all quantitative social science studies, but it is amplified in interaction models. In hierarchical regression analyses, interaction terms are typically modelled as the product of two observed predictor variables (i.e., main or direct effects), and the product's association with the outcome variable reflects how the main effects vary as a function of each other (Aguinis & Gottfredson, 2010). Because product terms have reduced variance, respective interactions tend to have small effect sizes in the prediction of complex traits (Duncan & Keller, 2011). We hasten to add that effect sizes being small renders them neither unimportant nor invalid: Small effects are the indispensable foundation upon which reliable and reproducible cumulative psychological science is built (Götz, Gosling, & Rentfrow, 2022). While improving psychological measurement may go some way in helping to identify interactions, its effects on statistical power are modest by comparison to the influence of sample sizes (Baranger, 2019; Cohen, 1992). It follows that increasing sample sizes is the key for advancing studies of developmental gene-environment interactions.

6. Bigger samples for better science

In his 1962 review of statistical power – or lack thereof – in psychological studies, Cohen (1962, p. 151) wrote "the answer is simple: increase sample size." We argue that this conclusion still holds today, but we recognise that achieving larger samples is challenging in developmental science, because collecting data from infants, toddlers, and children is expensive, complicated, and time consuming (Bergmann et al., 2018; Byers-Heinlein, Bergmann, & Savale, 2021; Davis-Kean & Ellis, 2019). We describe three ways that can help reduce the burden of collecting data from large childhood samples.

A first recommendation is for individual researchers, research teams, and laboratories to create data collection protocols that enable harmomnising, pooling, and sharing data across sites and data collection efforts for large-scale analyses. An example for a successful, ongoing collaboration effort of this kind – albeit not genetically sensitive – is the ManyBabies project (https://manybabies.github.io) that brings together investigators from more than 40 nations on 6 continents, with the aim to replicate and share best practices for developmental psychology research.

A second recommendation is to harmonise data from existing studies and pool them into larger datasets. This practice is common in GWA studies, where big consortia include hundreds of researchers and organisations to integrate data across biobanks, population cohort studies, and independent samples. These large-scale linkage collaborations have cumulated for example in the Psychiatric Genomics Consortium (https://www.med.unc.edu/pgc/) that focuses on mental health issues, and the Social Science Genetic Association Consortium (https://www.thessgac.org/) that targets social science outcomes, as its name suggests, among many other consortia.

Other examples of maximising sample sizes come directly from child developmental psychology studies. Databases such as CHILDES, PhonBank, and Homebank (MacWhinney, 1996; Rose & MacWhinney, 2014; van Dam et al., 2016) make audio recordings of children's speech and their naturalistic language environments accessible to a wide range of researchers. Similar database initiatives continue to emerge, for example Wordbank (http://wordbank.stanford.edu), an open database of children's vocabulary growth assessed by parents' reports, which currently holds data from over 85,000 children across 38 languages and 78 instruments. Because these databases do not (yet) include genotype data, they currently do not support conducting gene-environment interaction studies. However, they exemplify how to achieve bigger samples with greater statistical power in developmental science.

A third recommendation that cuts straight to testing gene-environment interactions is applying secondary data analyses to existing samples for whom genotype data are available, as well as measures of environmental conditions and phenotypes of interest. For example, Add Health (Highland, Avery, Duan, Li, & Harris, 2018) and the Wisconsin Longitudinal Study (Herd, Carr, & Roan, 2014) collected such data from overall 30,000 Americans, while the Twins Early

Development Study (~13,000 families with twins born in the mid-90s; Rimfeld et al., 2019) and the Millennium Cohort Study (~19,000 children born in 2000-2001; Fitzsimons et al., 2020) are UK population cohort studies for whom rich phenotype and genotype data are available.

To determine the sample sizes that might be needed to detect gene-environment interactions when using polygenic scores, we performed a series of Monte Carlo power analyses InteractionPoweR by Baranger et al. (2021) with 1000 simulations, rounding up to the nearest 100 participants (Fig. 2). We specified the main effects and the correlation between predictors to be 0.10, reflecting the effect sizes that are typically reported in psychological science (Götz et al., 2022). We specified the interaction terms' effect sizes to range from $\beta = 0.01$ and $\beta = 0.10$, which are small in absolute terms (Cohen, 1992) but large relative to the main effects (Aguinis et al., 2005; Götz et al., 2022), with increments of $\beta = 0.005$. Our simulations suggested that for detecting the smallest specified interaction effect size of $\beta = 0.01$ with statistical power of 80%, a sample of $N = 75{,}100$ was required (cf. Domingue, Trejo, Armstrong-Carter, & Tucker-Drob, 2020). Detecting the largest specified interaction effect size of $\beta = 0.10$ required a sample of $N = 800$ for 80% power, and $N = 3800$ were required for the mid-point interaction effect size of $\beta = 0.05$. Even though the curve for the required sample size for 80% power flattened substantially as the interaction effect sizes increased, our analyses suggest that samples of at least several thousand participants are needed to reliably identify gene-environment interactions in polygenic score models. This sample size demand will be exacerbated if multiple interactions between polygenic scores and a broad range of environmental factors are modelled simultanously (von Stumm et al., 2023).

7. Conclusion

Recent years have witnessed vast methodological advances for studying the gene-environment interplay of childhood cognitive development. Notwithstanding the remarkable research progress, a lingering issue in identifying gene-environment interactions is statistical power, because extremely large samples are needed to detect interaction effect sizes that are likely very small. Collecting data from large samples is always demanding and daunting, but particularly so in developmental science where infants, toddlers, and children are assessed. Yet, we propose that close collaborations and large-scale data sharing efforts can usher a new era of replicable gene-environment interaction findings. Although increasing sample sizes may seem costly, its returns will be far greater than continuing to publish interaction findings from small samples that are more often 'failures to replicate' than not.

Author contributions

SvS developed the concept of this review and led on the writing. AN edited and commented on the article.

CRediT authorship contribution statement

Sophie von Stumm: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Allie F. Nancarrow:** Writing – review & editing.

Declaration of competing interest

We have no conflicts of interest to declare.

Data availability

No data was used for the research described in the article.

Acknowledgments

This work was supported by a Nuffield Foundation project grant (EDO/44110). During the writing of this article, SvS was also supported by a British Academy Mid-career Fellowship (2022), a Jacobs Foundation CRISP Fellowship (2022-2027), and a fellowship from Paris Institue of Advanced Studies (2023-2024). We thank Megan Wright and Kirsty Wilding for their comments on previous versions of this paper.

References

Abraham, W. T., & Russell, D. W. (2008). Statistical power analysis in psychological research. *Personality Psychology Compass, 2*(1), 283–301. https://doi.org/10.1111/j.1751-9004.2007.00052.x

Aguinis, H., Beaty, J. C., Boik, R. J., & Pierce, C. A. (2005). Effect size and power in assessing moderating effects of categorical variables using multiple regression: A 30-year review. *Journal of Applied Psychology, 90*(1), 94–107. https://doi.org/10.1037/0021-9010.90.1.94

Aguinis, H., & Gottfredson, R. K. (2010). Best-practice recommendations for estimating interaction effects using moderated multiple regression. *Journal of Organizational Behavior, 31*(6), 776–786. https://doi.org/10.1002/job.686

Asbury, K., McBride, T., & Rimfeld, K. (2021). *Genetics and early intervention: Exploring ethical and policy questions*. Early Education Foundation. https://www.eif.org.uk/report/genetics-and-early-intervention-exploring-ethical-and-policy-questions.

Baranger, D. (2019, August 6). Interaction analyses - How large a sample do I need?. https://davidbaranger.com/2019/08/06/interaction-analyses-how-large-a-sample-do-i-need-part-3/.

Baranger, D. A., Finsaas, M., Goldstein, B., Vize, C., Lynam, D., & Olino, T. (2021). Tutorial: Power analyses for interaction effects in cross-sectional regressions. *PsyArXiv*, (August 4)https://doi.org/10.31234/osf.io/5ptd7

Bartels, M., van Beijsterveldt, C. E., & Boomsma, D. I. (2009). Breastfeeding, maternal education and cognitive function: A prospective study in twins. *Behavior Genetics, 39*(6), 616–622. https://doi.org/10.1007/s10519-009-9293-9

Belsky, J., & Pluess, M. (2009). Beyond diathesis stress: Differential susceptibility to environmental influences. *Psychological Bulletin, 135*(6), 885–908. https://doi.org/10.1037/a0017376

Bergmann, C., Tsuji, S., Piccinini, P. E., Lewis, M. L., Braginsky, M., Frank, M. C., & Cristia, A. (2018). Promoting replicability in developmental research through meta-analyses: Insights from language acquisition research. *Child Development, 89*(6), 1996–2009. https://doi.org/10.1111/cdev.13079

Boardman, J. D., Daw, J., & Freese, J. (2013). Defining the environment in gene-environment research: Lessons from social epidemiology. *American Journal of Public Health, 103*(Suppl. 1), S64–S72. https://doi.org/10.2105/AJPH.2013.301355

Bronfenbrenner, U., & Morris, P. A. (2006). The Bioecological Model of Human Development. In R. M. Lerner, & W. Damon (Eds.), *Handbook of child psychology: Theoretical models of human development* (6th ed., pp. 793–828). John Wiley & Sons, Inc.

Button, K. S., Ioannidis, J., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*(5), 365–376. https://doi.org/10.1038/nrn3475

Byers-Heinlein, K., Bergmann, C., & Savalei, V. (2021). Six solutions for more reliable infant research. *Infant and Child Development.* , Article e2296. https://doi.org/10.1002/icd.2296. Advance online publication.

Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., … Poulton, R. (2003). Influence of life stress on depression: Moderation by a polymorphism in the 5-HTT gene. *Science, 301*(5631), 386–389. https://doi.org/10.1126/science.1083968

Cattan, S., Fitzsimons, E., Goodman, A., Phimister, A., Ploubidis, G.B., & Wertz, J. (2022). Early childhood and inequalities. IFS Deaton Review of Inequalities.London: The IFS. Available at: https://ifs.org.uk/publications/early-childhood-inequalities-0.

Chabris, C. F., Lee, J. J., Cesarini, D., Benjamin, D. J., & Laibson, D. I. (2015). The fourth law of behavior genetics. *Current Directions in Psychological Science, 24*(4), 304–312. https://doi.org/10.1177/0963721415580430

Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *The Journal of Abnormal and Social Psychology, 65*(3), 145–153. https://doi.org/10.1037/h0045186

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. *Developmental psychopathology. Life-span human development* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. https://doi.org/10.1037//0033-2909.112.1.155

van Dam, M., Warlaumont, A. S., Bergelson, E., Cristia, A., Soderstrom, M., De Palma, P., & MacWhinney, B. (2016). HomeBank: An online repository of daylong child-centered audio recordings. *Seminars in Speech and Language, 37*(2), 128–142. https://doi.org/10.1055/s-0036-1580745

Davis-Kean, P. E., & Ellis, A. (2019). An overview of issues in infant and developmental research for the creation of robust and replicable science. *Infant Behavior and Development, 57*, Article 101339. https://doi.org/10.1016/j.infbeh.2019.101339

Deary, I. J. (2012). Intelligence. *Annual Review of Psychology, 63*, 453–482.

Demange, P. A., Malanchini, M., Mallard, T. T., Biroli, P., Cox, S. R., Grotzinger, A. D., … Nivard, M. G. (2021). Investigating the genetic architecture of noncognitive skills using GWAS-by-subtraction. *Nature Genetics, 53*(1), 35–44. https://doi.org/10.1038/s41588-020-00754-2

Dick, D. M., Agrawal, A., Keller, M. C., Adkins, A., Aliev, F., Monroe, S., … Sher, K. J. (2015). Candidate gene–environment interaction research: Reflections and recommendations. *Perspectives on Psychological Science, 10*(1), 37–59. https://doi.org/10.1177/1745691614556682

Domingue, B. W., Trejo, S., Armstrong-Carter, E., & Tucker-Drob, E. M. (2020). Interactions between polygenic scores and environments: Methodological and conceptual challenges. *Sociological Science, 7*, 465–486. https://doi.org/10.15195/v7.a19

Duncan, L. E., & Keller, M. C. (2011). A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *American Journal of Psychiatry, 168*(10), 1041–1049. https://doi.org/10.1176/appi.ajp.2011.11020191

Eaves, L. J., & Jinks, J. L. (1972). Insignificance of evidence for differences in heritability of IQ between races and social classes. *Nature, 240*(5376), 84–88. https://doi.org/10.1038/240084a0

Figlio, D. N., Freese, J., Karbownik, K., & Roth, J. (2017). Socioeconomic status and genetic influences on cognitive development. *Proceedings of the National Academy of Sciences, 114*(51), 13441–13446. https://doi.org/10.1073/pnas.1708491114

Fitzsimons, E., Moulton, V., Hughes, D. A., Neaves, S., Ho, K., Hemani, G., … Ring, S. (2020). *Collection of DNA samples and genetic data at scale in the UK millennium cohort study. CLS Working Paper 2020/7*. London: UCL Centre for Longitudinal Studies.

Götz, F. M., Gosling, S. D., & Rentfrow, P. J. (2022). Small effects: The indispensable foundation for a cumulative psychological science. *Perspectives on Psychological Science, 17*(1), 205–215. https://doi.org/10.1177/1745691620984483

Grant, M. D., Kremen, W. S., Jacobson, K. C., Franz, C., Xian, H., Eisen, S. A., … Lyons, M. J. (2010). Does parental education have a moderating effect on the genetic and environmental influences of general cognitive ability in early adulthood? *Behavior Genetics, 40*(4), 438–446. https://doi.org/10.1007/s10519-010-9351-3

Greven, C. U., Lionetti, F., Booth, C., Aron, E. N., Fox, E., Schendan, H. E., … Homberg, J. (2019). Sensory processing sensitivity in the context of environmental sensitivity: A critical review and development of research agenda. *Neuroscience & Biobehavioral Reviews, 98*, 287–305. https://doi.org/10.1016/j.neubiorev.2019.01.009

Grotzinger, A. D., Rhemtulla, M., de Vlaming, R., Ritchie, S. J., Mallard, T. T., Hill, W. D., … Tucker-Drob, E. M. (2019). Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nature Human Behavior, 3*(5), 513–525. https://doi.org/10.1038/s41562-019-0566-x

Hanscombe, K. B., Trzaskowski, M., Haworth, C. M., Davis, O. S., Dale, P. S., & Plomin, R. (2012). Socioeconomic status (SES) and children's intelligence (IQ): In a UK-representative sample SES moderates the environmental, not genetic, effect on IQ. *PLoS One, 7*(2), Article e30320. https://doi.org/10.1371/journal.pone.0030320

Herd, P., Carr, D., & Roan, C. (2014). Cohort profile: Wisconsin longitudinal study (WLS). *International Journal of Epidemiology, 43*(1), 34–41. https://doi.org/10.1093/ije/dys194

Highland, H. M., Avery, C. L., Duan, Q., Li, Y., & Harris, K. M. (2018). *Quality control analysis of add health GWAS data*. Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC. https://addhealth.cpc.unc.edu/wp-content/uploads/docs/user_guides/AH_GWAS_QC.pdf.

Kirkpatrick, R. M., McGue, M., & Iacono, W. G. (2015). Replication of a gene-environment interaction via multimodel inference: Additive-genetic variance in adolescents' general cognitive ability increases with family-of-origin socioeconomic status. *Behavior Genetics, 45*(2), 200–214. https://doi.org/10.1007/s10519-014-9698-y

Krapohl, E., Hannigan, L. J., Pingault, J. B., Patel, H., Kadeva, N., Curtis, C., … Plomin, R. (2017). Widespread covariation of early environmental exposures and trait-associated polygenic variation. *Proceedings of the National Academy of Sciences of the United States of America, 114*(44), 11727–11732. https://doi.org/10.1073/pnas.1707178114

MacWhinney, B. (1996). The CHILDES system. *American Journal of Speech-Language Pathology, 5*(1), 5–14. https://doi.org/10.1044/1058-0360.0501.05

Meehl, P. E. (1962). Schizotaxia, schizotypy, and schizophrenia. *American Psychologist, 17*(12), 827–838. https://doi.org/10.1037/h0041029

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist, 67*(2), 130. https://doi.org/10.1037/a0026699

Noble, C. L., Sala, G., Peter, M. S., Lingwood, J., Rowland, C. F., Gobet, F. R., & Pine, J. M. (2019). The impact of shared book reading on children's language skills: A meta-analysis. *Educational Research Review, 28*. https://doi.org/10.1016/j.edurev.2019.100290

Okbay, A., Wu, Y., Wang, N., Jayashankar, H., Bennett, M., Nehzati, S. M., … Young, A. I. (2022). Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nature Genetics, 54*(4), 437–449. https://doi.org/10.1038/s41588-022-01016-z

Plomin, R., & von Stumm, S. (2018). The new genetics of intelligence. *Nature Reviews Genetics, 19*(3), 148–159. https://doi.org/10.1038/nrg.2017.104

Pluess, M., & Belsky, J. (2013). Vantage sensitivity: Individual differences in response to positive experiences. *Psychological Bulletin, 139*(4), 901–916. https://doi.org/10.1037/a0030196

Protzko, J., Aronson, J., & Blair, C. (2013). How to make a Young child smarter: Evidence from the database of raising intelligence. *Perspectives on Psychological Science, 8*(1), 25–40. https://doi.org/10.1177/1745691612462585

Rimfeld, K., Malanchini, M., Spargo, T., Spickernell, G., Selzam, S., McMillan, A., … Plomin, R. (2019). Twins early development study: A genetically sensitive investigation into behavioral and cognitive development from infancy to emerging adulthood. *Twin Research and Human Genetics, 22*(6), 508–513. https://doi.org/10.1017/thg.2019.56

Rose, Y., & MacWhinney, B. (2014). The PhonBank Project: Data and software-assisted methods for the study of phonology and phonological development. In J. Durand, U. Gut, & G. Kristoffersen (Eds.), *The Oxford handbook of corpus phonology* (pp. 380–401). Oxford: Oxford University Press. NIHMS761002 https://psyling.talkbank.org/years/2014/phonbank.pdf.

Rowe, D. C., Jacobson, K. C., & Van den Oord, E. J. (1999). Genetic and environmental influences on vocabulary IQ: Parental education level as moderator. *Child Development, 70*, 1151–1162.

Scarr-Salapatek, S. (1971). Race, social class, and IQ. *Science, 174*, 1285–1295.

Sigelman, C. K., & Rider, E. A. (2009). The ecology of developmental processes. *Developmental psychopathology. Life-span human development* (6th ed., pp. 468–495). Belmont, CA: Wadsworth Cengage Learning.

van der Sluis, S., Dolan, C. V., Neale, M. C., & Posthuma, D. (2008). Power calculations using exact data simulation: A useful tool for genetic study designs. *Behavior Genetics, 38*, 202–211. https://doi.org/10.1007/s10519-007-9184-x

Sokolowski, H. M., & Ansari, D. (2018). Understanding the effects of education through the lens of biology. *npj Science of Learning, 3*(1), 1–10. https://doi.org/10.1038/s41539-018-0032-y

von Stumm, S., & d'Apice, K. (2022). From genome-wide to environment-wide: Capturing the environome. *Perspectives on Psychological Science, 17*(1), 30–40. https://doi.org/10.1177/1745691620979803

von Stumm, S., Kandaswamy, R., & Maxwell, J. (2023). Gene-environment interplay in early life cognitive development. *Intelligence, 98*. https://doi.org/10.1016/j.intell.2023.101748

Tucker-Drob, E. M., & Bates, T. C. (2016). Large cross-national differences in gene× socioeconomic status interaction on intelligence. *Psychological Science, 27*(2), 138–149. https://doi.org/10.1177/0956797615612727

Turkheimer, E., Haley, A., Waldron, M., d'Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science, 14*(6), 623–628. https://doi.org/10.1046/j.0956-7976.2003.psci_1475.x

Zuckerman, M. (1999). Diathesis-stress models. In I. M. Zuckerman (Ed.), *Vulnerability to psychopathology: A biosocial model* (pp. 3–23). American Psychological Association. https://doi.org/10.1037/10316-001.