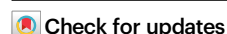


Neural–genetic–environmental evidence for a disease factor in mental and physical health multimorbidity

Received: 9 December 2024

Accepted: 7 October 2025

Published online: 18 November 2025



Jin Chen^{1,2,15}, Yuning Zhang^{3,4,15}, Shu Liu^{5,6}, Miguel Garcia-Argibay^{3,7,8}, Tianye Jia^{4,9}, Jujiao Kang^{4,9}, Wei Li^{1,2}, Marco Solmi^{10,11}, Hongyi Sun³, Wenqi Liu¹², Congying Chu¹³✉, Samuele Cortese^{3,14}✉ & Jiaojian Wang^{1,2}✉

Increasing evidence reveals the presence of multimorbidity across physical and mental disorders. A general disease factor (*d* factor) has been recently identified to capture the shared liability across these conditions, yet its biological basis remains poorly understood. Here, using data from the UK Biobank, we reveal the *d* factor's neural, genetic, and environmental underpinnings. We show that the *d* factor is associated with extreme negative deviations in grey matter volume and white matter microstructure. A genome-wide association study identifies its genetic loci and correlations with unhealthy lifestyle, anthropometric measures, and mood-related phenotypes. Furthermore, Mendelian randomization suggests a causal effect of living environmental deprivation on the *d* factor. Mediation analysis further reveals that the *d* factor links this environmental adversity to individual differences in brain structure. Our findings establish a multi-level biological characterization of general disease liability, connecting environmental, genetic, and neural factors and inform transdiagnostic approaches to prevention and treatment.

The co-occurrence of multiple mental and physical illnesses is prevalent and poses tremendous challenges in diagnosis and treatment¹. This body-mind association in medicine has been observed to be bidirectional—physical illnesses may contribute to the development of

mental disorders² and vice versa³. For instance, it has been reported that psychiatric and psychological factors play an important role in at least 30% of dermatologic disorders⁴, and patients with a range of somatic disorders are more alexithymic than controls⁵. Individuals

¹State Key Laboratory of Primate Biomedical Research, Institute of Primate Translational Medicine, Kunming University of Science and Technology, Kunming, China. ²Yunnan Key Laboratory of Primate Biomedical Research, Kunming, China. ³Developmental EPI (Evidence synthesis, Prediction, Implementation) lab, Centre for Innovation in Mental Health, School of Psychology, Faculty of Environmental and Life Sciences, University of Southampton, Southampton, UK. ⁴Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China. ⁵Key Laboratory of Genetic Evolution and Animal Models, National Research Facility for Phenotypic and Genetic Analysis of Model, Kunming, China. ⁶Animals (Primate Facility), National Resource Center for Non-Human Primates, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China. ⁷School of Medical Sciences, Faculty of Medicine and Health, Örebro University, Örebro, Sweden. ⁸Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. ⁹Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence (Fudan University), Ministry of Education, Shanghai, China. ¹⁰SCIENCES lab, Department of Psychiatry, University of Ottawa, Ontario, ON, Canada. ¹¹Regional Centre for the Treatment of Eating Disorders and On Track: The Champlain First Episode Psychosis Program, Department of Mental Health, The Ottawa Hospital, Ontario, ON, Canada. ¹²Data Science Research Center, Kunming University of Science and Technology, Kunming, China. ¹³Brainnetome Center & National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China. ¹⁴Clinical and Experimental Sciences (CNS and Psychiatry), Faculty of Medicine, University of Southampton, Southampton, UK. ¹⁵These authors contributed equally: Jin Chen, Yuning Zhang. ✉e-mail: chucongying@gmail.com; Samuele.Cortese@soton.ac.uk; jiaojianwang@uestc.edu.cn

suffering from comorbid psychiatric and physical illnesses typically experience more severe deficits in various cognitive and behavioral functions and suffer more than those with a single illness^{6,7}. A better understanding of the nature of mental and physical multimorbidity is urgently needed because it has not only important clinical significance but also important theoretical value in medicine, psychology and neuroscience⁸.

The use of latent concepts, such as the *p* factor⁹ and the more recently proposed neuropsychopathological (*NP*) factor¹⁰, to explain multimorbidity is common in psychiatric research. These factors, however, primarily address psychiatric multimorbidity. Recent findings indicate that mental-physical multimorbidity is especially prevalent among older adults¹¹, highlighting the need for a broader construct. To meet this need, a recent study introduced a latent general disease factor, referred to as the *d* factor (for disease), which serves as a common underlying construct for both physical and mental disorders¹². Specifically, unlike the *p* factor and *NP* factor which are limited to psychiatric domains, the *d* factor extends to incorporate both mental and physical health conditions, offering a comprehensive framework for understanding multimorbidity^{13,14}. This expansion from a mental-specific construct to a general disease liability underscores the necessity to investigate shared neurobiological, genetic, and environmental influences across diverse health domains, rather than limiting analyses to psychiatric or somatic conditions alone.

The brain, as a central organ of the human body, has been consistently implicated in both mental and physical illnesses. Accumulating evidence shows that structural and functional abnormalities in the brain, such as gray matter (GM) atrophy and white matter (WM) disruption are associated with various chronic diseases and psychiatric conditions^{15,16}. Similarly, genomic factors also contribute significantly to the etiology of these conditions. In particular, genome-wide association studies (GWAS) have identified numerous genetic variants linked to both mental disorders¹⁷ and physical illnesses¹⁸, providing important insights into the genetic architecture underlying disease susceptibility and multimorbidity. The conceptualization of transdiagnostic constructs such as the *p* factor, despite ongoing debate around their structural validity¹⁹, has been increasingly informed by genomic research²⁰. Moreover, integrating GWAS of the *d* factor with Mendelian randomization (MR) analyses allows us to test the directionality and causality of its associations with phenotypes, thereby advancing our understanding of the mechanisms underlying general health and disease vulnerability.

Evidence has repeatedly indicated a strong relationship between living environmental deprivation and brain maturity or development²¹. For example, Marshall et al. reported that children from lower-income families exhibited lower cognitive test scores, smaller cortical volume, and smaller cortical surface area with increasing risk of lead exposure²². Similarly, the negative impact of environmental risk exposure on both mental and physical illness etiology is well documented²³. For example, characteristics of neighborhoods with low socioeconomic status are associated with greater cardiovascular risk²⁴, and residential crowding is associated with psychological distress in the general population^{25,26}. These findings suggested that our environment shapes the developmental trajectories of behavior and the brain in terms of both structure and function. It remains unclear, however, whether disease processes mediate the relationship between environmental deprivation and brain health or whether alterations in brain structure/function mediate the relationship between environmental exposure and disease status.

Therefore, in the current study, we aimed to provide a comprehensive characterization of the *d* factor by integrating multimodal data from the UK Biobank. We first estimated individual *d* factor scores using the bifactor exploratory structural equation modeling (Bifactor-ESEM). Building on this construct, we examined its associations with individualized brain structural deviations using normative modeling, conducted a GWAS to identify associated genetic loci and genetic

correlations with healthy-related lifestyle phenotypes, anthropometric measures, and mood-related phenotypes, and further applied MR to test potential causal effects of environmental deprivation on the *d* factor. Mediation analyses were further employed to explore whether the *d* factor serves as a pathway linking environmental deprivation to brain structural alterations. By integrating *d* factor, brain phenotypes, genetic risk, and environmental deprivation into a unified analytical framework, our study provides a multi-level perspective on the biological mechanisms underlying general health vulnerability. An overview of the study design is provided in Fig. 1.

Results

d factor for mental and physical multimorbidities

Following our previous work¹², the unifactor confirmatory factor analysis (Unifactor-CFA), correlated factors confirmatory factor analysis (Corfactor-CFA), and bifactor confirmatory factor analysis (Bifactor-CFA) were fitted with the data from UK Biobank to define the optimal model characterizing mental and physical multimorbidities. We found that the bifactor model outperformed the other two models ($CFI = 0.974$, $TFI = 0.970$, $RMSEA = 0.006$). The fitting performances of the three models are presented in Supplementary Table 1, and the detailed loading values for each mental or physical illness derived from the three models are shown in Supplementary Fig. 1. Following the strong performance of the Bifactor-CFA model relative to the Unifactor-CFA and Corfactor-CFA models, we further evaluated the bifactor exploratory structural equation modeling (Bifactor-ESEM) since the Bifactor-ESEM allows for cross-loadings providing a more flexible structure that better accounts for the inherent multidimensionality in mental and physical disorders comorbidities²⁷. As shown in Supplementary Table 2 and Table 3, the Bifactor-ESEM model showed superiority in fitting and reliability metrics compared to the Bifactor-CFA model ($CFI = 0.981$, $TFI = 0.977$, $RMSEA = 0.006$), including higher Omega and OmegaH values across the *d*, mental, and physical factors, indicating better model reliability and fitting accuracy ($\Omega_{ESEM}(d) = 0.910$, $\Omega_{CFA}(d) = 0.617$). This suggests that Bifactor-ESEM provides a more accurate representation of the latent structure underlying mental and physical multimorbidities, making it a preferable choice for estimating the *d* factor in this study. Thus, a Bifactor-ESEM model was fitted to the baseline data of all subjects from the UK Biobank, yielding three distinct factors: the *d* factor, which captured the extent to which mental and physical illnesses co-occurred; the mental health (*mh*) factor, which reflected the general pattern of psychopathological symptoms, with cross-loadings onto physical health conditions; and the physical health (*ph*) factor, which captured the general pattern of physical illnesses, with cross-loadings onto mental health conditions. The detailed loading values for each mental or physical illness derived from the Bifactor-ESEM are shown in Supplementary Table 4.

To validate the effectiveness of the *d* factor from the Bifactor-ESEM model (Basic framework of the Bifactor-ESEM shown in Supplementary Fig. 2a), we conducted a linear regression analysis on a mental and physical multimorbidity sample ($n = 21,256$) to examine the relationship between the *d* factor and the number of diagnoses. A strong positive association ($R^2 = 0.76$) was found indicating that the *d* factor effectively quantified the level of mental-physical illness multimorbidities. That is with the number of diseases increase, the higher *d* factor score (Supplementary Fig. 2b).

Individual brain structural variation associated with the *d* factor

To investigate the neural basis of the *d* factor, we examined its associations with regional GM volume and WM microstructure. Normative models of GM volume and fractional anisotropy (FA) were constructed using data from healthy participants to delineate age-related trajectories of brain structural features. To reveal the neural basis of *d* factor, we applied multiple linear regression analyses to investigate the

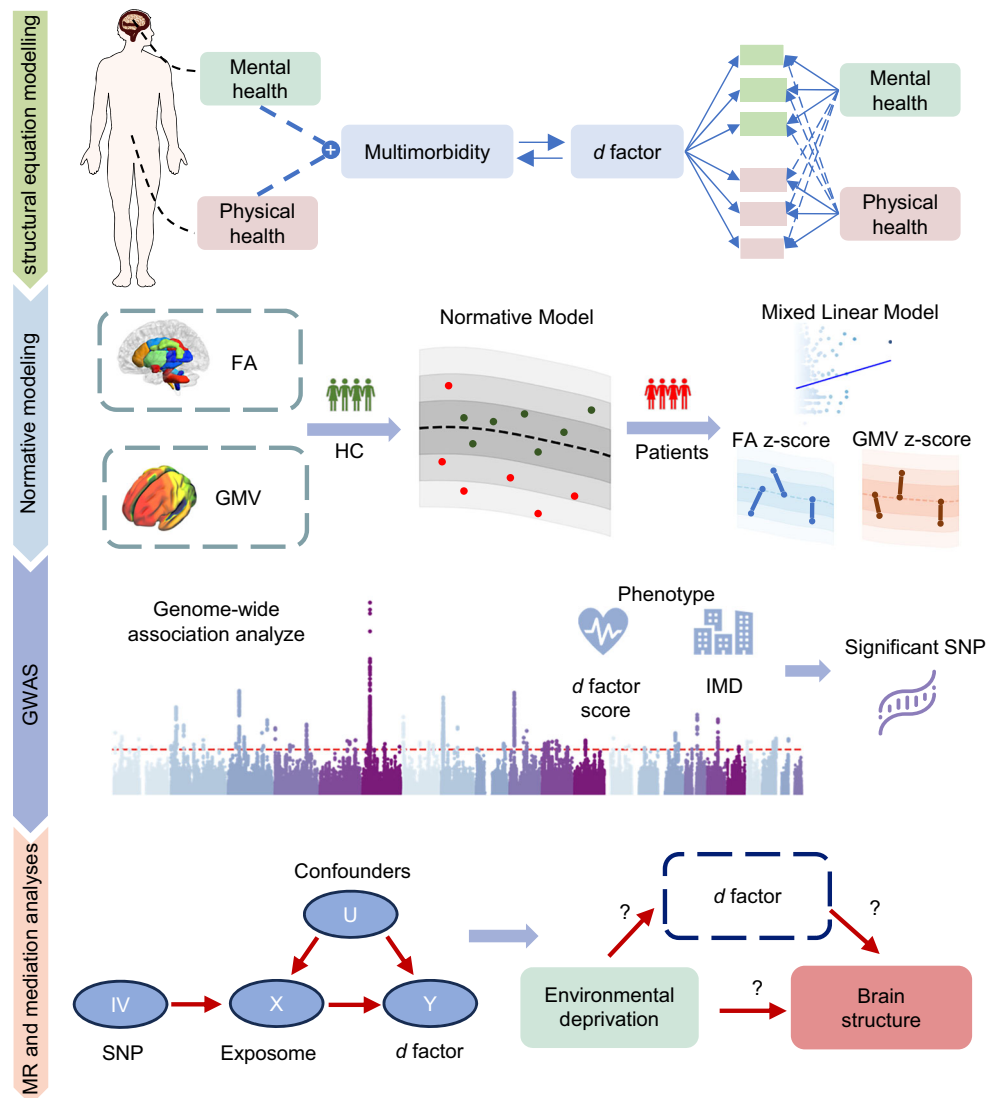


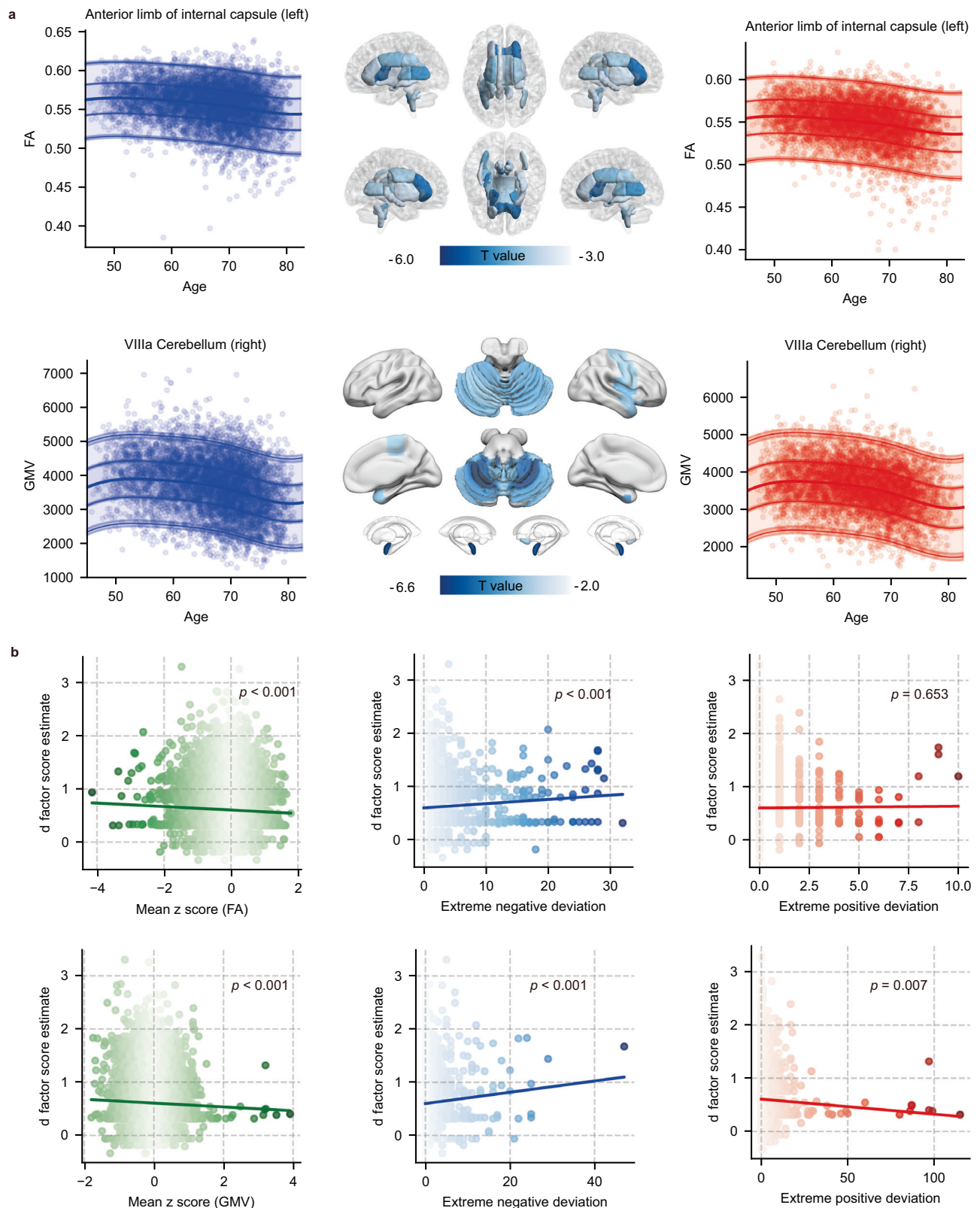
Fig. 1 | The analytical pipelines used in this study. This study developed a general disease (d) to quantify the physical and mental multimorbidities using bifactor exploratory structural equation modeling (Bifactor-ESEM) and investigated its neural and genetic basis and associations with environmental exposure. We first used the Bifactor-ESEM to obtain the d factor and corresponding mental and physical factors. Then, we constructed normative models for white matter (WM) microstructure and regional gray matter (GM) volume to further reveal the neural basis of d factor. We depicted the trends of brain structural indicators for each brain region or WM tract with age and sex. The relationship between d factor score estimates and mean deviation z-scores, number of extreme deviations of brain

structural indicators was investigated. Additionally, we conducted genome-wide association analyses (GWAS) to identify loci associated with d factor and the index of multiple deprivation (IMD) and performed enrichment analysis for genes associated with d factor. Mendelian randomization (MR) analysis was used to reveal a potential causal relationship between IMD and the d factor. The mediation analyses were finally used to explore the potential relationship between d factor, environmental exposure, i.e., IMD, and changes of brain structures. Image of human body provided by Servier Medical Art (<https://smart.servier.com/>), licensed under CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/>).

associations between the d factor and the z-scores of GM volume and FA across brain regions and WM tracts, respectively. The normative aging trajectories of a brain region and a WM tract with the largest statistic t values in multiple linear regression analysis were shown in males and females (Fig. 2a). The majority of GM regions and WM tracts exhibited negative associations with the d factor. Specifically, the d factor showed significant associations with the cerebellum (such as lobule VIIIa) and amygdala, as well as in WM tracts including the anterior limb of the internal capsule, posterior thalamic radiation, and superior cerebellar peduncle after multiple comparison correction (Fig. 2a, t values from the multiple linear regression analyses are presented in Supplementary Data 1–2). To further quantify regional structural deviations in individuals with higher disease burden, we calculated the mean deviation z-scores for GM volume and FA of each brain region and each tract in participants with multimorbidity

(Supplementary Fig. 3). Consistent with the regression results, the participants with multimorbidity showed markedly lower mean z-scores in the cerebellum, amygdala, posterior thalamic radiation, and superior cerebellar peduncle.

To further investigate the relationships between deviations and d factor scores, we calculated the individual-specific mean z-scores for both gray and white matter in patients. We then used the mean z-scores and extreme deviations ($|z| > 2.6$, with the definition provided in the Methods) for gray and white matter as independent variables, and the d factor score estimates as the dependent variable, to construct mixed linear models (Fig. 2b). We found that the individual-specific mean FA z-scores and the individual-specific mean GM volume z-scores both had significant negative effects on d factor score estimates ($\beta = -0.033$, $p < 0.001$; $\beta = -0.036$, $p < 0.001$). The extreme negative deviation in FA had a significant positive effect on d factor



score estimates ($\beta = 0.008$, $p < 0.001$), while extreme positive deviations in FA had no significant effect ($\beta = 0.003$, $p = 0.653$). Additionally, extreme negative deviations in GM volume had a significant positive effect on d factor score estimates ($\beta = 0.011$, $p < 0.001$), whereas extreme positive deviations in GM volume had a significant negative effect ($\beta = -0.003$, $p = 0.007$). These findings suggested that more extreme negative deviations in both WM tracts and regional GM

volume were associated with higher d factor scores, indicating that mental and physical multimorbidity disrupt WM microstructure and GM volume.

Genome-wide association study of d factor score

To reveal the genetic basis of d factor, we conducted GWAS of d factor. To account for potential biases in the genetic association analysis, we

Fig. 2 | Normative models for WM microstructure (fractional anisotropy: FA) and gray matter (GM) volume and their relationships with *d* factor score estimates. **a** For each sex, normative models of FA in WM tracts and GM volume in brain regions were separately constructed. Only the *t* values of regions that survived Bonferroni correction ($p < 0.05$) in the multiple linear regression analyses between regional *z*-scores and the *d* factor scores are shown. Age-related trajectories are plotted for the WM tract and brain region with the largest absolute *t* values in males and females, respectively (red = female, blue = male). Individual patient data points are plotted as semi-transparent scatter points. The central line represents the median predicted value from the healthy normative model, and shaded bands indicate the 25th–75th and 5th–95th percentile ranges. Wider bands correspond to increased epistemic uncertainty. **b** Left panel: mixed linear model analyses revealed significantly negative associations between *d* factor score estimates and the mean *z*-score of FA in WM tracts ($\beta = -0.033$, $p = 4.08 \times 10^{-7}$, 97.5% CI: [-0.045, -0.020]) and between *d* factor score estimates and the mean *z*-score of GM ($\beta = -0.036$, $p = 1.80 \times 10^{-5}$, 97.5% CI: [-0.052, -0.020]). Middle panel: the *d* factor score estimate showed significantly positive associations with the number of extreme negative deviations of FA ($\beta = 0.008$, $p = 1.75 \times 10^{-5}$, 97.5% CI: [0.004, 0.012]) and GM volume ($\beta = 0.011$, $p = 2.96 \times 10^{-5}$, 97.5% CI: [0.006, 0.016]). Right panel: mixed linear model analyses were used to identify the relationships between *d* factor score estimates and extreme positive deviation of FA and GM volume. There are no significant associations between *d* factor score estimates and extreme positive deviation of FA ($\beta = 0.003$, $p = 0.653$, 97.5% CI: [-0.011, 0.017]) while there are significantly negative associations between *d* factor score estimates and extreme positive deviation of GM volume ($\beta = -0.003$, $p = 0.007$, 97.5% CI: [-0.005, -0.001]). The extreme deviation is defined as extreme deviation $|z| > 2.6$. All tests are two-sided.

examined the genomic inflation factor. There was an observed inflation in the genomic inflation factor at the quantile-quantile (QQ) plot ($\lambda = 1.517$; Fig. 3a). Although this λ value suggests some degree of inflation to induce potential confounding, our linkage disequilibrium score regression (LDSC) analysis results support the robustness of these findings. Specifically, the LDSC intercept was estimated at 1.085 (SE = 0.010), which is close to the ideal value of 1, indicating that the majority of inflation observed in λ is likely attributable to true polygenic signal rather than confounding factors. Additionally, the ratio metric was 0.134 (SE = 0.016), underscoring that most of the inflation in the test statistics originates from genuine polygenic effects rather than population structure or other confounding influences. Thus, despite the elevated λ , the LDSC results justify the reliability of the association signals, minimizing the confounding of our significant loci findings. We identified a total of 30 genomic risk loci (33 lead SNPs) that met the genome-wide significance threshold ($p < 5 \times 10^{-8}$) (Fig. 3a and Supplementary Data 3).

Phenotype-based gene set enrichment

To elucidate the broader phenotypic associations and biological functions of the genes linked to the *d* factor score, we first mapped 71 genes from genome-wide significant SNPs (Supplementary Data 4). These genes were subsequently used for phenotypic enrichment analysis against the GWAS Catalog within the Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA) platform. Significance was defined using a Benjamini-Hochberg false discovery rate (FDR) threshold of $p < 0.05$. As illustrated in Fig. 3b, the most significantly enriched phenotypes spanned several domains, including lifestyle behaviors, cognitive traits, sleep patterns, and dietary habits. We next performed Gene Ontology biological process (GO-BP) enrichment analysis on these 71 genes using the ToppGene Suite. Of these, 58 genes were successfully annotated in ToppGene and retained for analysis. After correcting for multiple comparisons with the Benjamini-Hochberg FDR method ($p < 0.05$), several biological processes were significantly enriched. These were primarily associated with sensory and visual system development, smooth muscle tissue development, and telomere- and calcium-related regulatory mechanisms (Fig. 3c).

Genetic correlations between the *d* factor and anthropometric and lifestyle traits

To further investigate whether the *d* factor is genetically associated with lifestyle-related phenotypes, anthropometric measures, and mood-related phenotypes, we selected 34 commonly studied phenotypes, including body composition indices and lifestyle behaviors such as tobacco exposure, alcohol intake, and sleep patterns, and performed bivariate LDSC analysis. After correction for multiple testing using the Benjamini-Hochberg false discovery rate (FDR < 0.05) method, 32 out of the 34 traits showed significant genetic correlations with the *d* factor (Fig. 3d and Supplementary Table 8).

Among lifestyle-related phenotypes, household tobacco exposure ($r_g = 0.66$), the presence of smokers in the household ($r_g = 0.54$), insomnia ($r_g = 0.48$) and alcohol intake ($r_g = 0.43$), were all strongly positively correlated with the *d* factor. Notably, past tobacco smoking, a lower scores indicating greater smoking exposure, showed a negative correlation ($r_g = -0.30$). These patterns suggest that genetic susceptibility to poor health is intertwined with predispositions toward adverse lifestyle behaviors. Additionally, a lower average total household income was strongly associated with higher *d* factor scores ($r_g = -0.54$). Anthropometric phenotypes showed similarly robust genetic overlap, body mass index showed the strongest genetic correlation with the *d* factor ($r_g = 0.54$), followed by diastolic blood pressure ($r_g = 0.44$), and weight ($r_g = 0.40$), highlighting shared genetic influences on general health and metabolic status. Other body composition measures, including systolic blood pressure ($r_g = 0.38$) and pulse rate ($r_g = 0.18$), also correlated significantly but more modestly. Hand grip strength was negatively correlated ($r_g = -0.23$ for the right hand; $r_g = -0.22$ for the left hand), indicating that lower genetically determined muscular strength is linked to higher *d* factor scores. Finally, mood-related phenotypes, including mood swings ($r_g = 0.49$), loneliness ($r_g = 0.46$), and fed-up feelings ($r_g = 0.46$), showed significant positive genetic correlations, underscoring the broad psychosocial dimension of the *d* factor.

Genetic evidence for the link between index of multiple deprivation and *d* factor

Given that environmental exposure is a high-risk contributor to both mental and physical diseases, we conducted a bidirectional two-sample MR analysis to investigate the causal relationship between socioeconomic deprivation and the *d* factor. Genome-wide association summary statistics for the index of multiple deprivation (IMD) were obtained from a sample of 168,475 individuals, and the GWAS results for the *d* factor were derived from a non-overlapping sample of 168,475 individuals. A total of 16 independent SNPs were selected as instrumental variables for IMD after linkage disequilibrium clumping and removal of variants with known pleiotropic effects. In the forward direction, MR analysis revealed a significant causal effect of IMD on the *d* factor, suggesting that socioeconomic deprivation contributes to individual differences in disease accumulation (Fig. 4a). Sensitivity analyses using the weighted median and MR-Egger methods yielded consistent results. There was no evidence of horizontal pleiotropy, as indicated by a non-significant MR-Egger intercept (intercept = 0.0007, $p = 0.902$). In contrast, the reverse MR analysis did not support a causal effect of the *d* factor on IMD, and this null result remained consistent across all MR methods. Again, no evidence of horizontal pleiotropy was detected (Supplementary Fig. 4). Taken together, these findings support a unidirectional causal relationship in which socioeconomic deprivation influences the *d* factor, rather than the reverse.

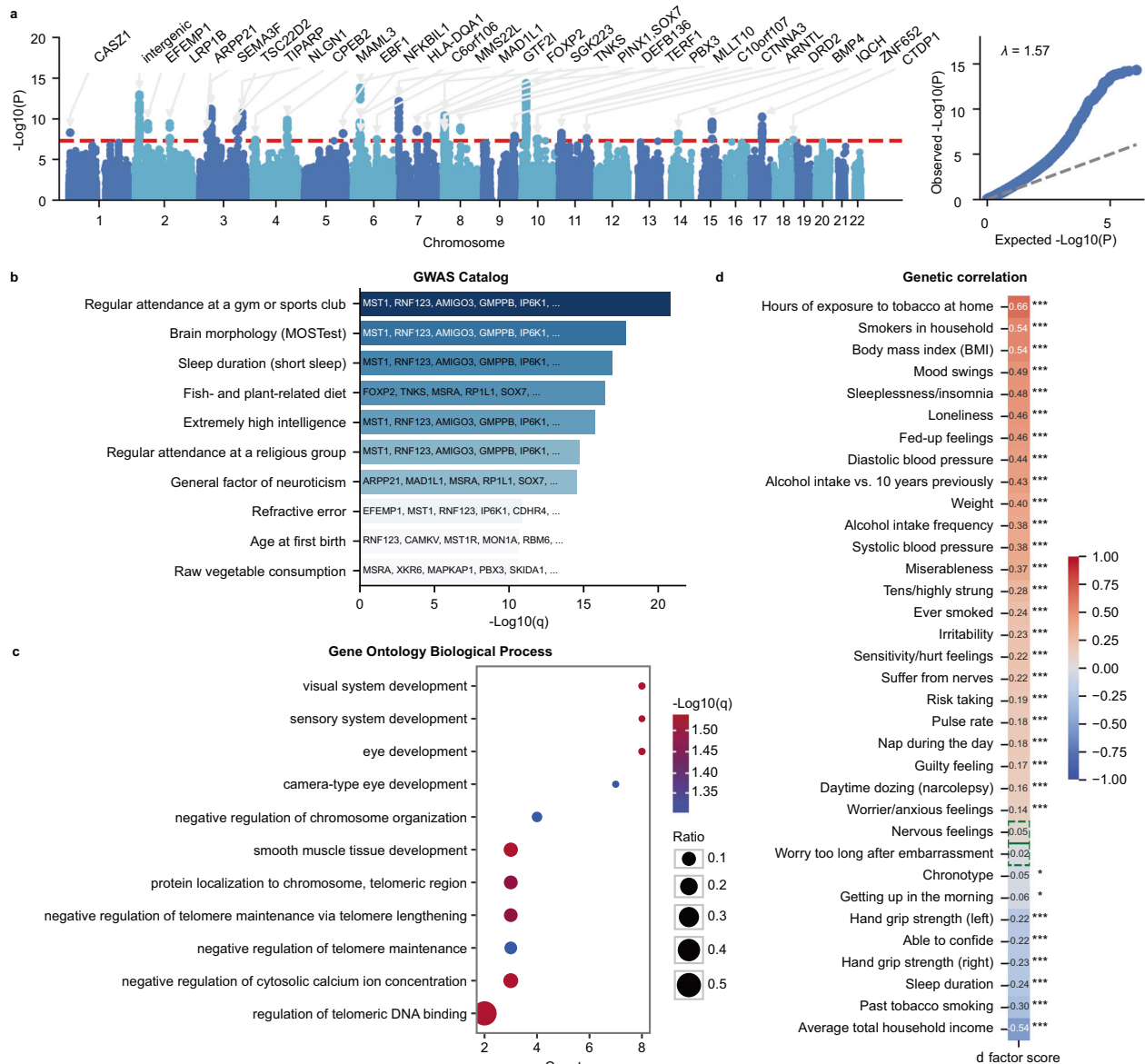


Fig. 3 | Genome-wide association analysis of d factor scores estimates. **a** The Manhattan plot displays SNP associations across the genome. Each point represents the $-\log_{10}(p)$ value from a linear mixed model association of the d factor with individual SNPs. The genome-wide significance threshold is set at 5×10^{-8} to correct for multiple comparisons. The QQ plot ($\lambda = 1.52$) shows the observed versus expected p value distribution, indicating modest inflation of association signals. All tests are two-sided, and multiple testing is controlled using the genome-wide significance threshold. **b** GWAS catalog enrichment analysis revealed previously reported phenotypes sharing genetic associations with the significant SNPs of the d

factor identified in this study, with the top 10 most significantly associated traits highlighted. **c** GO biological pathway enrichment analysis illustrates the biological processes significantly associated with the genes linked to the identified SNPs after FDR-BH correction ($q > 0.05$). **d** Heatmap of genetic correlations between the d factor score and phenotypic phenotypes, showing 34 phenotypes ranked by the absolute r_g value. The green dashed line indicates non-significance after FDR-BH correction ($q > 0.05$). Significance thresholds are denoted as *** $q < 0.001$, ** $q < 0.01$, and * $q < 0.05$.

Results of the mediation analyses

To further examine the relationships among IMD, the d factor, and brain structures of GM volume and WM microstructure, we performed mediation analyses using IMD as the independent variable, the d factor or brain structures as mediators. Significant mediation effects of IMD on both WM and GM were observed only when the d factor was specified as the mediator (Fig. 4b and Supplementary Fig. 5). For WM tracts, we found that the d factor significantly mediated the effect of IMD on the microstructure of WM tracts in the brain in patients with mental and physical multimorbidities ($ab/c' = 40\%$, $ab = -2 \times 10^{-4}$, 95% CI, $[-3 \times 10^{-4}, -1 \times 10^{-4}]$) (Fig. 4b). Given that the direct effect of IMD score on the microstructure of WM tracts was not significant

($c' = -5 \times 10^{-4}$, 95% CI, $[-1 \times 10^{-3}, 5 \times 10^{-4}]$), thus this result represents full mediation effect. In addition, we found that the d factor also had a partial mediating effect ($ab/c' = 3.1\%$, $ab = -1 \times 10^{-4}$, 95% CI, $[-2 \times 10^{-4}, -1 \times 10^{-4}]$) on the relationship between IMD score and brain regional GM volume in patients with mental and physical multimorbidities because of the significant direct effect ($c' = -3 \times 10^{-3}$, 95% CI, $[-4 \times 10^{-3}, -2 \times 10^{-3}]$) (Fig. 4b).

Discussion

This study builds on our previous work¹ on the general disease (d) factor by integrating brain structures, environmental deprivations, and genetic factors to uncover its neural and genetic underpinnings.

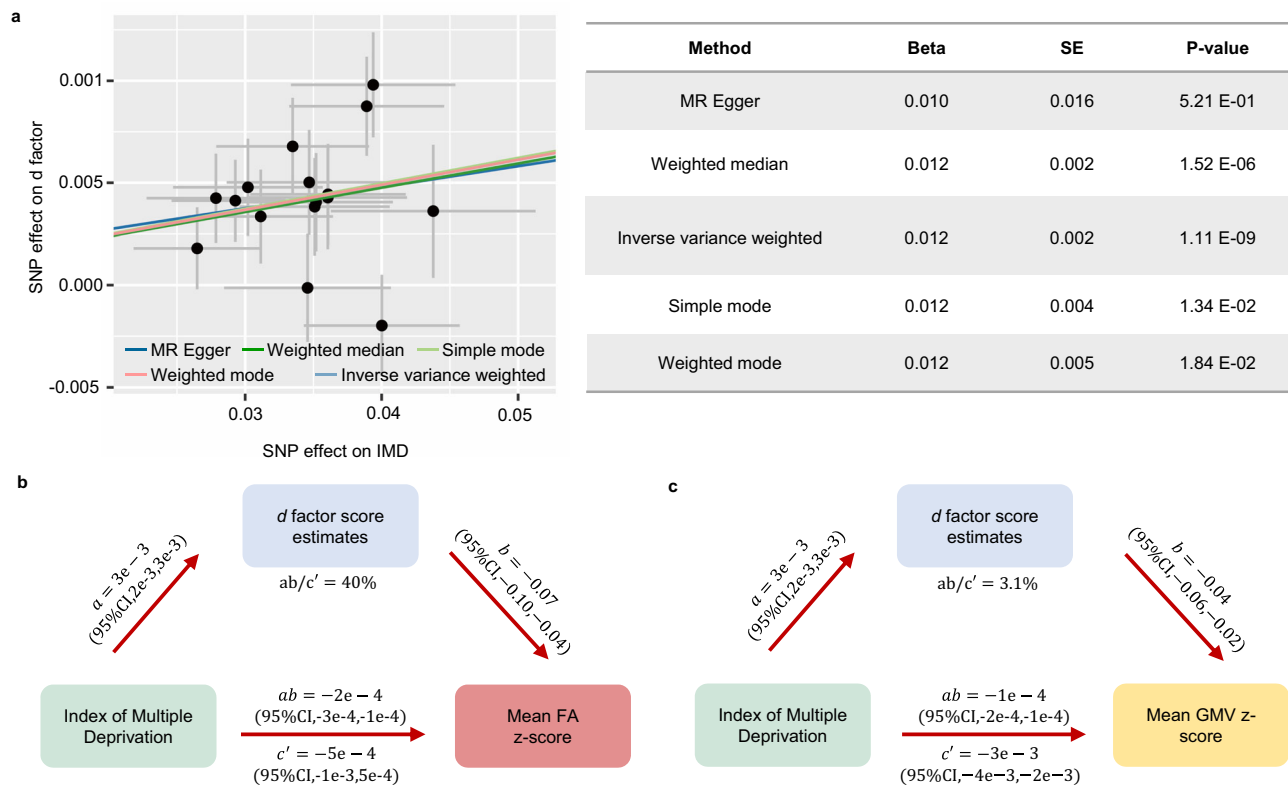


Fig. 4 | Mendelian randomization analysis of index of multiple deprivation (IMD) and d factor and mediating effect of the d factor score estimates on the influence of environmental exposures on the brain structures. a Mendelian randomization analysis was performed to assess the relationship between the IMD ($n = 168,475$) and the d factor ($n = 168,475$). Causal effect estimates were derived using weighted median, inverse variance weighted (IVW), simple mode, weighted median, and weighted mode, with separate lines showing the fitted causal effect for each method. Each point represents the estimated effect (β) of an individual genetic variant (SNP) on the outcome, with the center indicating the point estimate

and error bars showing the 95% confidence intervals. **b** The IMD was mediated by the d factor score estimates for effect on WM microstructure ($ab = -2 \times 10^{-4}$, 95% CI, $[-3 \times 10^{-4}, -1 \times 10^{-4}]$). The direct effect of the IMD on WM microstructure suggested that ($c' = -5 \times 10^{-4}$, 95% CI, $[-1 \times 10^{-3}, 5 \times 10^{-4}]$), irrespective of the degree of multimorbidity, the IMD was not significantly associated with the mean FA z-score. Thus, the IMD achieved a full mediating effect on the brain through the d factor score estimates. **c** The IMD was mediated by the d factor score estimates for effects on GM volume ($ab = -1 \times 10^{-4}$; 95% CI, $[-2 \times 10^{-4}, -1 \times 10^{-4}]$). The IMD achieved a partial mediating effect on the GM volume through d factor score estimates.

Leveraging data from over 500,000 individuals, we extended our previous model into a bifactor-ESEM framework to more robustly estimate d factor scores¹³. By establishing normative models of brain structures, we found that d factor scores were positively associated with the number of brain regions exhibiting extreme negative deviations, primarily within the sensorimotor and limbic circuits and their associated WM tracts. GWAS of the d factor scores indicated that the associated genes were enriched for neurodevelopmental and sensory processes and showed significant genetic correlations with unhealthy lifestyle phenotypes and anthropometric measurements, highlighting the genetic basis of the d factor. Mendelian randomization further supported a potential causal effect of environmental deprivation (i.e., IMD) on the d factor, which in turn mediated the influence of environmental deprivation on brain structures. By uncovering the neural and genetic basis of the d factor, these findings provide a foundation for identifying shared mechanisms across disorders and developing transdiagnostic prevention and intervention strategies.

In our previous study, we used a Bifactor-CFA approach to estimate the d factor, but this model cannot achieve cross-loadings to account for the interplay between mental and physical health. To strengthen previous results, in the present study we optimized the d factor estimate using a Bifactor-ESEM model¹³, which permits cross-loadings while preserving a bifactor structure. This approach provided superior fit, higher reliability, and a more balanced allocation of

general and specific variance compared with Bifactor-CFA. The better performance reflects a strong general disease factor, while also revealing meaningful variance in the specific mental and physical factors, underscoring the multidimensional nature of multimorbidity. The optimized model exhibited the robustness of the d factor while also revealing meaningful domain-specific variance, offering a more realistic characterization of multimorbidity.

By deriving the d factor from clinical diagnosis data to capture the emergent covariance patterns among mental and physical health indicators, we conceptualize the d factor as reflecting shared vulnerabilities shaped by both common environmental exposures and the dynamic interplay between internal and external factors. For example, genetic predisposition, early-life adversity, and chronic stress may together exert an overarching susceptibility to poor health, thereby influencing both mental and physical outcome²⁸. Likewise, socioeconomic deprivation and accumulated environmental stressors can shape health trajectory^{29,30}, contributing to the observed covariance structure. Therefore, the d factor emerges as a manifestation of complex biopsychosocial processes rather than a simple sum of various diseases or associations with brain structure or function. The theory of d factor is supported by prior research on health and multimorbidity^{31,32}, which increasingly recognizes that shared variation across diverse disease states may arise from common environmental and structural determinants rather than solely from person-internal vulnerabilities.

To further elucidate the shared structural determinants, we examined the neuroanatomical correlates of the estimated d factor. The analysis revealed associations with structural characteristics of the brain, particularly implicating the cerebellum and its connected WM tracts. The cerebellum is not only central to motor coordination but also plays an increasingly recognized role in cognitive and affective regulation³³. Its structure and function are highly sensitive to environmental exposures³⁴, including environmental profiles³⁵ and lifestyle-related stressors, which may accelerate vulnerability to multimorbidity. Notably, WM microstructural integrity and GM volume are known to decline with age^{36,37}, but multimorbidity may exacerbate this trajectory. Individuals with higher multimorbidity burden are frequently affected by chronic conditions associated with neuropathological changes, including neuronal loss, synaptic degeneration, inflammation, oxidative stress, and vascular dysfunction, and these factors accelerate age-related deterioration of WM integrity and GM volume and eventually lead to brain atrophy^{38–40}. These findings highlight a shared neurobiological pathway through which aging and multimorbidity may reinforce each other, contributing to widespread structural brain atrophy.

GWAS analyses of the d factor identified significant genetic loci. The associated genes were enriched for functions related to brain morphology and lifestyle factors, consistent with our brain imaging results showing that d factor scores were associated with brain structure alter. Functional enrichment analyses further highlighted pathways in neurodevelopment and sensory processing, aligning with literature reporting altered brain and sensory function in multimorbidity^{15,41}. Notably, enrichment was also observed in telomere- and calcium-related processes, which have been linked to cellular aging, stress reactivity, and systemic regulation in both mental and physical health⁴². These findings are consistent with our imaging evidence and point to the possibility that multimorbidity accelerates age-related brain decline, with the d factor capturing a broader aging-related vulnerability. Genetic correlation analyses revealed that the d factor was strongly associated with health-related traits, including tobacco exposure, body mass index, blood pressure, sleep characteristics, and alcohol intake. These associations are consistent with previous reports linking each of these factors to both physical and mental disorders^{43–46}, suggesting that common liabilities may be shaped through pathways of cardiometabolic risk, health behaviors, and circadian regulation. With these findings, we suggest that the d factor captures not only shared genetic susceptibility but also its convergence with aging processes and lifestyle pathways that jointly influence disease risk. This perspective moves beyond single-disease explanations and positions the d factor as a unifying construct that integrates genetic, neurobiological, and environmental influences on multimorbidity.

Environmental exposure is an established risk factor for mental and physical illness^{35,47}. In this study, we used the IMD to represent multidimensional environmental adversity and, through MR, identified a positive causal effect of IMD on the d factor, supporting the view that the d factor reflects a broad biopsychosocial mechanism of risk accumulation. Furthermore, mediation analyses indicated that the d factor fully mediates the effect of environmental deprivation on GM volume and partially mediates its effect on white matter microstructure. Together with our genetic findings showing enrichment in aging-related pathways, lifestyle- and income-related traits, these results emphasize that genetic predispositions and environmental adversity converge on the d factor to shape brain structure and confer transdiagnostic risk. Our findings align with existing literature on environmental influences on the brain^{22,48,49} and health^{35,47}, and underscore the theoretical relevance of the d factor as a unifying construct environmental exposure, genetic liability, neurobiology, and clinical multimorbidity. The partial mediation observed in WM suggests its particular sensitivity to environmental and genetic interplay,

potentially offering a target for mechanistic investigation into shared pathways across mental and physical disorders.

Although we made substantial efforts to address potential biases, several limitations should be acknowledged when interpreting the results. First, the UK Biobank cohort is not fully representative of the general UK population. Participants tend to be healthier, better educated, and less socioeconomically deprived than the population average, which may limit the generalizability of our findings to underrepresented groups⁵⁰. Second, our reliance on inpatient hospital records may introduce selection bias, as these data predominantly capture individuals with more severe or chronic conditions requiring hospitalization. As a result, associations between mental and physical health conditions may be overestimated compared to those observed in the general population, particularly among individuals with milder symptoms who do not require inpatient care⁵¹. Nevertheless, inpatient records provide clinically verified diagnoses and are less prone to recall or social desirability biases than self-reported data, thereby enhancing the reliability and objectivity of disease phenotyping. Therefore, the d factor is not fully independent of time, which largely reflects a cumulative disease burden based on hospital diagnoses.

In conclusion, our findings carry significance across several levels. Conceptually, they suggest the existence of a shared underlying factor that influences both mental and physical health. This challenges a common misunderstanding that the d factor is simply another label for multimorbidity. Although both frameworks address the co-occurrence of multiple health conditions, they differ markedly in methodological basis and interpretation. Multimorbidity typically refers to the presence of two or more chronic illnesses in one person, often quantified through counts or simple pairwise associations. In contrast, the d factor represents a latent dimension that explains shared variance across a wide spectrum of health conditions, uncovered through more sophisticated statistical approaches such as bifactor modeling¹³. In this sense, our work advances a way of conceptualizing health. From a research perspective, studies on transdiagnostic risk and mechanisms of disease transmission may benefit from integrating both mental and physical conditions. This raises an important question: should mental and physical disorders continue to be treated as distinct categories, or is it more appropriate to view them together under the broader framework of health conditions¹²? Furthermore, the d factor framework may provide a foundation for creating advanced tools to assess risk and guide transdiagnostic interventions that address shared underlying mechanisms. In terms of prevention, it points to possible etiopathophysiological mechanisms that can be targeted to jointly prevent mental and physical conditions, which would strengthen the value and scope of preventive efforts. Notably, current clinical guidelines in mental health rarely provide guidance on managing associated physical conditions, highlighting the need for more comprehensive protocols that encompass both domains. From a policy and service organization standpoint, our results provide evidence in favor of health policies that encourage the integration of mental and physical health services, which could enhance resource efficiency and improve overall health outcomes. Furthermore, they support an integrated care approach in Consultation-Liaison psychiatry for old populations, emphasizing the importance of mental health expertise within general medical settings and, conversely, the inclusion of medical considerations in mental health care. This is particularly relevant given the shared underlying vulnerability (d factor) linking physical and mental conditions commonly encountered in Consultation-Liaison practice. Overall, in terms of service organization, our results strongly discourage the siloed approach to care that is common in certain mental healthcare systems (e.g., in the UK), where services are fragmented and disconnected.

Methods

Participants

This study was performed in accordance with the Declaration of Helsinki. Informed consent was obtained from all participants prior to

their inclusion in the study. The UK Biobank data were approved by the Northwest Multi-Center Research Ethics Committee (MREC) (REC reference 21/NW/0157). All the data used in the current study are from the UK Biobank, which is a large biomedical database containing genetic, imaging, behavioral, and health information from 502,505 participants (229,122 males, 45.6%; 273,383 females, 54.4%) aged 38–73 years living in the UK. Information on sex was based on self-reported biological sex (Field ID: 31), with no genetic verification performed. Multimodal brain imaging data were collected in 2014–2020 from three mirrored imaging centers in Manchester, Reading, and Newcastle⁵². Our confirmatory factor analysis to construct the *d* factor score estimates included a total of 502,505 participants, of which 35,056 participants' structural and diffusion MRI data (16,390 males, 46.8%; 18,666 females, 53.2%) were used to examine the neural basis of the *d* factor. The mediation models included data from 10,383 participants (5461 males, 52.6%; 4922 females, 47.4%) due to missing data patterns across key variables of the *d* factor, environmental deprivation, and the brain. The demographic information of these participants is summarized in Supplementary Table 5.

Materials/measurements

Mental and physical illnesses used to construct the *d* factor. The participants' diagnosis status for 30 illnesses was used to construct the bifactor CFA models. Participants' hospital inpatient records were coded according to the International Classification of Disease version 10 (ICD-10)⁵³ (field ID: 41270) and then binarized for analysis. The major mental disorders included conduct disorders, depression, anxiety, phobia, pervasive developmental disorders, obsessive-compulsive disorder, schizophrenia, bipolar disorder, psychosis, mania, and adjustment disorder. The physical conditions included stroke, asthma, dermatitis, arthritis, dorsalgia, ulcers, colitis, irritable bowel syndrome, kidney ureter stones, hearing loss, visual disturbances, obesity, insulin-dependent diabetes mellitus, non-insulin-dependent diabetes mellitus, essential (primary) hypertension, secondary hypertension, hernia, bronchitis, and movement disorders. The criteria for selecting mental and physical illnesses were based on a previously published study¹². The details for all the selected mental and physical illnesses are given in Supplementary Tables 6, 7.

Brain phenotypes. We utilized multimodal brain imaging data, encompassing both brain T1 structural MR images and diffusion MR images, to identify the associations between the *d* factor and the regional GM volume and microstructure of WM tracts. Data acquisition was conducted using a standard Siemens Skyra 3 T scanner with a standard 32-channel RF receiver head coil. The preprocessed region-wise GM volume and tract-wise FA data were provided by UK Biobank. Here, we described the key preprocessing steps related to the generated brain features. Specifically, to assess regional GM volume, UK Biobank employed the FAST algorithm in the FSL toolkit (<https://fsl.fmrib.ox.ac.uk/fsl>). Using FAST, the GM volume map was generated for each participant. Then, the whole brain was segmented into 139 regions of interest (ROIs) using the Harvard-Oxford cortical and sub-cortical atlases⁵⁴ and the Diedrichsen cerebellar atlas⁵⁵. To facilitate data analysis, the mean GM volume of each ROI was obtained from the UK Biobank. For WM tracts, we selected mean fractional anisotropy (FA) values to characterize WM microstructure. To avoid partial volume effects and registration errors, the whole-brain FA map of each subject was first aligned onto a standardized WM skeleton. Then, the whole-brain WM was segmented into 48 tracts using the ICBM-DTI-81 WM atlas and overlaid onto the FA map on the skeleton⁵⁶. The mean FA value of each WM tract on the skeleton was calculated for each participant for further analyses. With the preprocessed structural and diffusion MRI data, the mean GM volume of 139 brain areas and the mean FA of 48 WM tracts for each subject were used to explore the neural basis of the *d* factor.

Environmental deprivation. Because the IMD has a strong relationship with health in both rural and urban areas⁵⁷, we examined the relationships among environmental risk factors characterized using the IMD, brain structures, and the *d* factor. The IMD offers a more comprehensive and detailed description of deprivation based on more factors than the Townsend index. As IMD scores are calculated in different ways in England, Scotland and Wales and the participants with multimorbidities from England accounted for approximately 94% (10,383/11,044) of the total number of patients, while very few participants were from Scotland and Wales, the mediation analyses for environmental deprivation were only conducted for patients from England. A total of 8 items, including the IMD, crime, education, employment, health, housing, income, and living environment, were provided for these participants. The 8 item scores were obtained from the English IMD original data (<https://www.gov.uk/government/collections/english-indices-of-deprivation>). Among these 8 items, the IMD differed from the other 7 items in that the IMD score is an overall score of the relative degree of multiple deprivations experienced by each community.

Genotypes, quality control and PCA. Genotype data for UK Biobank participants were generated using the UK BiLEVE or UK Biobank Axiom arrays, and imputed to the Haplotype Reference Consortium (HRC), UK10K, and 1000 Genomes reference panels⁵⁸. Quality control (QC) was performed at both the sample and variant level. At the sample level, we restricted analyses to individuals of genetically inferred European ancestry, excluding those with sex chromosome aneuploidy, discordant sex information, or excessive relatedness (third-degree relatives or closer). In addition, we retained only individuals included in the phasing of autosomes (chromosomes 1–22). At the SNP level, we excluded variants with minor allele frequency (MAF) < 0.005, Hardy–Weinberg equilibrium $p < 1 \times 10^{-6}$, genotype missingness > 0.05, or not present in the HapMap3 reference panel, ensuring high-quality and cross-cohort comparable markers. After QC, a total of 336,950 participants (156,024 males, 52.6%; 180,926 females, 47.4%) and 1,153,470 SNPs were retained. Population structure was assessed using principal component analysis (PCA). To minimize bias from linkage disequilibrium (LD), SNPs were pruned to retain approximately independent markers with pairwise LD $r^2 < 0.2$, resulting in 138,037 SNPs. PCA was then performed on these SNPs across all retained individuals.

Statistical analysis

Confirmatory factor analysis and *d* factor. Following our previous work¹², a Bifactor-CFA model was constructed, and the bifactor score estimates (*d* factor) were used for further analysis. Three validated *d* factor models were examined for the enrolled mental and physical disorders based on the entire UK Biobank sample ($n = 502,505$) using confirmatory factor analysis to obtain the optimal model characterizing mental and physical multimorbidities⁵⁹.

The first model tested was the Cor-factor model, which assumes that all illnesses, both mental and physical, are correlated with each other. The second model is the Uni-factor model, which posits that all illnesses could be best accounted for by a single factor. The last model is the bifactor model, which postulates that mental and physical illnesses would be best characterized by distinct individual factors, while an overarching disease dimension (that is, the *d* factor in this study) could provide the most comprehensive explanation for the observed data.

The model fit was evaluated using chi-square values, the comparative fit index (CFI), the Tucker-Lewis index (TLI), and the root mean square error of approximation (RMSEA). The lower the RMSEA is, the better the model fit is (< 0.06 = good model fit); the higher the CFI and TLI are, the better the model fit is (> 0.95 = good model fit).

Bifactor-ESEM. To enhance our approach and address the limitations of CFA modeling, we also implemented a bifactor exploratory structural equation model. Although preliminary comparisons indicated that the Bifactor-CFA model performed well, we pursued Bifactor-ESEM to evaluate its advantages over Bifactor-CFA, with the goal of selecting the optimal model for estimating the d factor. Bifactor-ESEM is an advanced model that combines exploratory and confirmatory techniques, allowing for a more thorough assessment of multidimensional constructs by estimating cross-loadings across non-target factors⁶⁰. This method offers a more comprehensive understanding of factor structures, as it relaxes the constraints of CFA by permitting all factors, not just the hypothesized primary factor, to contribute to item variance. This estimation process has been shown to reduce factor loadings and interfactor correlations, resulting in factorial solutions that more accurately reflect the complexity of multidimensional data.

The model fit was evaluated using the Explained Common Variance (ECV)^{61,62}, Omega^{63,64}, and Omega Hierarchical (OmegaH)⁶⁵, along with the Factor Determinacy (FD)⁶⁶ and H index⁶⁷. Higher ECV values indicate a strong general factor, while elevated omega values suggest a highly reliable multidimensional composite⁶⁸. Specifically, omegaH represents the proportion of systematic variance in unit-weighted total scores attributed to individual differences on the general factor; when omegaH is high (e.g., > 0.8), total scores can be considered essentially unidimensional. Factor score estimates should only be considered reliable when FD values exceed 0.9⁶⁹. Furthermore, high H values (e.g., > 0.8) indicate a well-defined latent variable, suggesting greater stability across studies.

Normative modeling. Since univariate statistical analysis tests only group effects while ignoring individual-level differences, we thus employed a normative model to obtain individual-specific FA and GM volume deviation maps^{70–75}. The normative model utilized data from a large cohort of healthy individuals ($n = 24,014$) and included demographic information such as age, sex and scanning site to estimate the mean and variance (referred to as the normative range) of the response variables. In this study, we employed a Bayesian linear regression (BLR) model using the Predictive Clinical Neuroscience toolkit (PCNtoolkit) (<https://pcntoolkit.readthedocs.io/en/latest/>) to estimate normative models of brain structure. Specifically, we modeled the volume of each gray matter region and the fractional anisotropy (FA) of each white matter tract. Sex and scanning site were included as covariates. To account for nonlinear effects of age, we applied a B-spline basis expansion over age. Additionally, we used likelihood warping to address potential non-Gaussianity in the data⁷⁶. Model performance was evaluated using tenfold cross-validation to ensure the generalizability of the results. This Bayesian approach computes probability distributions over the linear coefficients that define the relationships between the variables, updating these distributions based on the observed data. By specifying priors over the coefficients, this method yields robust and unbiased estimates of the underlying relationships, while also accounting for uncertainty in regions with fewer data points. Subsequently, the FA of each WM tract and the GM volume of each brain region for individuals with mental and physical multimorbidities were predicted using the fitted model.

To quantify the individual deviation, the deviation of each WM tract or brain region derived from the normative model was calculated for each subject with mental and physical multimorbidity. The deviation was defined as the average FA of each fiber tract or mean GM volume of each region compared to the predicted FA or GM volume using normative models. Then, the deviation was transformed to a z score, which measures the degree to which a true value deviates from the normative model-predicted value⁷². The z score of

the deviation was calculated using the following formula:

$$z_{ij} = \frac{y_{ij} - \hat{y}_{ij}}{\sqrt{\sigma_{ij}^2 + \sigma_{nj}^2}} \quad (1)$$

Where $\hat{y}_{ij}$ represents the predicted FA value, y_{ij} is the actual FA value, σ_{ij} is the predicted variance, and σ_{nj} is the normative model variance.

Due to the intrinsic differences between individuals, we used extreme deviation to characterize the abnormal subjects who showed larger deviations from the expected deviations of the normative model. To define extreme deviations, a threshold value of $|z| > 2.6$ corresponding to unilateral $p < 0.005$ was used to identify brain structures (brain regions or WM tracts) with extreme deviations compared to the model predictions. Subsequently, we calculated the number of brain structures showing positive or negative extreme deviations for each participant in the comorbid samples.

Finally, to assess whether these extreme individual differences could predict multimorbidities, we separately constructed mixed linear models for positive and negative extreme deviations with the extreme deviations as independent variables and d factor score estimates as dependent variables, and $p < 0.05$ was considered to indicate statistical significance⁷⁷. The model was estimated using the restricted maximum likelihood method.

Multivariable regression analysis. To examine associations between the d factor score and regional brain measures, we conducted separate multivariable linear regression analyses for each brain region. For each GM region ($n = 139$) and WM tract ($n = 48$), the d factor score was entered as the dependent variable, and the regional z -value was the independent variable of interest. Sex and age were included as covariates, and head size was additionally controlled for in the GM volume models. Statistical significance was defined as Bonferroni-corrected $p < 0.05$ across all regions and tracts.

Genome-wide association analysis. To investigate the genetic contributions to the d factor, we conducted a genome-wide association study (GWAS) using genome-wide complex trait analysis (GCTA). The analysis was performed with GCTA version 1.92.3 beta3⁷⁸. Individuals with non-missing phenotype and covariate data were included. For estimating genetic variance, the Haseman-Elston regression method was applied⁷⁹. GWAS was performed using the mixed-model fastGWA approach⁸⁰, adjusting for age, sex, and the first four genetic principal components to account for population structure. Model parameters were tuned using 1000 randomly selected null SNPs. All computations were conducted using 10 threads to improve computational efficiency.

Linkage disequilibrium score regression. To assess the genetic architecture of the d factor and evaluate potential confounding, we applied linkage disequilibrium score regression (LDSC). Specifically, LDSC was used to examine the quantile-quantile (QQ) plot and genomic inflation, distinguishing true polygenic signal from inflation due to population structure or other confounders. Given that the genomic inflation factor (λ) can increase with sample size even in the absence of bias⁸¹, we focused on the LDSC intercept and ratio metrics, which provide more accurate estimates of confounding-free polygenic signal. SNP-based heritability of the d factor was also estimated using LDSC.

Gene-set and ontology analyses. Following the GWAS, functional annotation and gene set enrichment analyses were performed using the Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA)⁸². Independent genome-wide significant SNPs from the GWAS ($p < 5 \times 10^{-8}$) were mapped to genes using default positional

mapping strategy of FUMA, which assigns each SNP to the nearest protein-coding gene within a ± 10 kb window. To identify phenotypes sharing genetic associations with our results, the list of mapped genes was subsequently tested for enrichment against the NHGRI-EBI GWAS Catalog⁸³ using the tool integrated within FUMA. Enrichment was calculated based on a hypergeometric test, with multiple testing correction applied using the Benjamini-Hochberg FDR method. Phenotypes with an FDR-adjusted p -value < 0.05 were considered significantly enriched. Additionally, Gene Ontology (GO) enrichment analysis was performed on the list of mapped genes using the ToppGene Suite⁸⁴. The analysis was conducted against the default background of all human protein-coding genes. default settings of ToppGene were used, and significance was determined after Benjamini-Hochberg FDR correction (FDR < 0.05).

Genetic correlations analysis. To evaluate the genetic correlations between the d factor score and a wide range of phenotypes, we applied bivariate LDSC⁸⁵. Based on previous literature, we selected 34 phenotypes related to anthropometric measures, health-related lifestyle, and mood (Supplementary Table 8), including but not limited to: (1) body mass index (BMI)⁴⁵, a key predictor of overall health and disease risk; (2) smoking behavior^{43,44}; (3) alcohol consumption⁸⁶; (4) sleep duration⁴⁶; and (5) mental health status⁸⁷. For these 35 behavioral and health-related traits, we used GWAS summary statistics reported by Jiang et al.^{80,88}, derived from genome-wide genotyping data of 456,422 individuals of European ancestry in the UK Biobank. LDSC leverages the correlation structure of SNPs due to linkage disequilibrium to estimate the shared genetic architecture between two traits and is robust to sample overlap. Multiple testing was corrected using the FDR, with a significance threshold of $p < 0.05$.

Mendelian randomization. To assess the causal relationship between IMD and the d factor, we conducted a mendelian Randomization (MR) analysis using genome-wide SNPs significantly associated with IMD as instrumental variables (IVs). To avoid bias due to participant overlap, separate GWAS for d factor ($N = 168,475$) and IMD ($N = 168,475$) were performed using two different and non-overlapping groups of participants. Age, sex and the first 4 genetic PCs were adjusted in all GWAS analyses. Linkage disequilibrium clumping ($r^2 < 0.2$) was performed with the European 1000 Genomes Project phase 3 as the reference, using a 100 kb clumping window and a threshold $p < 5 \times 10^{-8}$. All MR analyses were conducted using the TwoSampleMR package in R.

Mediation analyses for the d factor, brain structure, and environmental risk factors. To explore the interactions among the brain structure, environmental risk factors, and disease, we performed mediation analyses among d factor score estimates, scores of the IMD in England (that is, environmental deprivation), and the mean z score of FA across all WM tracts or z score of GM volume across all brain regions for subjects with multimorbidities⁸⁹ (Fig. 1). Here, we did not perform mediation analysis for each WM tract or brain region but rather calculated an overall index to characterize a general brain abnormality in patients with mental and physical multimorbidities.

We then constructed different mediation models in which the d factor score estimate, mean GM volume z -score, and mean FA z -score were employed as mediator variables. The IMD served as independent variables, while the mean FA z -score, mean GM volume z -score, and d factor score estimates were the dependent variables. Notably, mediation effects were present in all models, with only the d factor score estimate serving as a mediator variable to test the mediating relationship of multimorbidity scores between multiple poverty measure scores and WM and GM changes. For mediation analysis, sex and age were added as covariates to each path of the mediation analysis model. The

following equation was used to describe the mediation interactions:

$$M = aX + \varepsilon_2 \quad (2)$$

$$Y = c'X + bM + \varepsilon_3 \quad (3)$$

where X refers to the independent variable, Y refers to the dependent variable, M is the mediating variable, c' refers to the direct effect, and the product ab is the indirect effect.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

This research was conducted using data from the UK Biobank under access application number 197910. UK Biobank makes its data available to all bona fide researchers for any health-related research that serves the public interest, without granting preferential or exclusive access to any individual or group. All researchers are subject to the same application process and approval criteria, as defined by UK Biobank. For further details regarding the access procedure, please visit the UK Biobank website: <http://www.ukbiobank.ac.uk/register-apply/>. Source data are provided with this paper.

Code availability

The code used in this study is available at GitHub (<https://github.com/Chenjin1110/d-factor-code>) and is citable via Zenodo <https://doi.org/10.5281/zenodo.17280644>⁹⁰.

References

1. Daré, L. O. et al. Co-morbidities of mental disorders and chronic physical diseases in developing and emerging countries: A meta-analysis. *BMC Public Health* **19**, 304 (2019).
2. Loh, H. H., Lim, L. L., Yee, A. & Loh, H. S. Association between subclinical hypothyroidism and depression: An updated systematic review and meta-analysis. *BMC Psychiatry* **19**, 12 (2019).
3. De Hert, M. et al. Physical illness in patients with severe mental disorders. I. Prevalence, impact of medications and disparities in health care. *World Psychiatry* **10**, 52–77 (2011).
4. Gupta, M. A. & Gupta, A. K. Psychiatric and psychological comorbidity in patients with dermatologic disorders: Epidemiology and management. *Am. J. Clin. Dermatol.* **4**, 833–842 (2003).
5. Sifneos, P. E. The prevalence of 'alexithymic' characteristics in psychosomatic patients. *Psychother. Psychosom.* **22**, 255–262 (1973).
6. Fiorillo, A. & Sartorius, N. Mortality gap and physical comorbidity of people with severe mental disorders: the public health scandal. *Ann. Gen. Psychiatry* **20**, 52 (2021).
7. Steffen, A., Nübel, J., Jacobi, F., Bätzing, J. & Holstiege, J. Mental and somatic comorbidity of depression: a comprehensive cross-sectional analysis of 202 diagnosis groups using German nationwide ambulatory claims data. *BMC Psychiatry* **20**, 142 (2020).
8. Arrondo, G. et al. Associations between mental and physical conditions in children and adolescents: An umbrella review. *Neurosci. Biobehav. Rev.* **137**, 104662 (2022).
9. Caspi, A. et al. The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clin. Psychol. Sci.* **2**, 119–137 (2014).
10. Xie, C. et al. A shared neural basis underlying psychiatric comorbidity. *Nat. Med.* **29**, 1232–1242 (2023).
11. van den Brink, A. M., Gerritsen, D. L., Voshaar, R. C. O. & Koopmans, R. T. Residents with mental–physical multimorbidity living in long-

- term care facilities: prevalence and characteristics. A systematic review. *Int. Psychogeriatr.* **25**, 531–548 (2013).
12. Brandt, V. et al. First evidence of a general disease (“d”) factor, a common factor underlying physical and mental illness. *World Psychiatry* **22**, 335–337 (2023).
 13. Garcia-Argibay, M. et al. General disease factor: evidence of a unifying dimension across mental and physical illness in children and adolescents. *BMJ Ment. Health* **28**, e301592 (2025).
 14. Sun, H. et al. Large-scale evidence of a general disease (“d”) factor accounting for both mental and physical health disorders in different age groups. *Psychol. Med.* **55**, e78 (2025).
 15. Shang, X. et al. Association of a wide range of individual chronic diseases and their multimorbidity with brain volumes in the UK Biobank: A cross-sectional study. *EClinicalMedicine* **47**, 101413 (2022).
 16. Lee, A. et al. Comparison of fornix and stria terminalis connectivity among first-episode schizophrenia, chronic schizophrenia and healthy controls. *Korean J. Biol. Psychiatry* **26**, 8–13 (2019).
 17. Mullins, N. et al. GWAS of suicide attempt in psychiatric disorders and association with major depression polygenic risk scores. *Am. J. Psychiatry* **176**, 651–660 (2019).
 18. Wells, H. R. R. et al. GWAS identifies 44 independent associated genomic loci for self-reported adult hearing difficulty in UK biobank. *Am. J. Hum. Genet.* **105**, 788–802 (2019).
 19. Waldman, I. D. et al. Recommendations for adjudicating among alternative structural models of psychopathology. *Clin. Psychol. Sci.* **11**, 616–640 (2023).
 20. Williams, C. M. et al. Characterizing the phenotypic and genetic structure of psychopathology in UK Biobank. *Nat. Ment. Health* **2**, 960–974 (2024).
 21. Noble, K. G., Norman, M. F. & Farah, M. J. Neurocognitive correlates of socioeconomic status in kindergarten children. *Dev. Sci.* **8**, 74–87 (2005).
 22. Marshall, A. T. et al. Association of lead-exposure risk and family income with childhood brain outcomes. *Nat. Med.* **26**, 91–97 (2020).
 23. Evans, G. W. & Kantrowitz, E. Socioeconomic status and health: The potential role of environmental risk exposure. *Annu. Rev. Public Health* **23**, 303–331 (2002).
 24. Smith, G. D., Hart, C., Watt, G., Hole, D. & Hawthorne, V. Individual social class, area-based deprivation, cardiovascular disease risk factors, and mortality: the Renfrew and Paisley Study. *J. Epidemiol. Community Health* **52**, 399–405 (1998).
 25. Evans, G. W., Palsane, M. N., Lepore, S. J. & Martin, J. Residential density and psychological health: the mediating effects of social support. *J. Pers. Soc. Psychol.* **57**, 994 (1989).
 26. Cohen, S., Evans, G. W., Stokols, D. & Krantz, D. S. *Behavior, Health, and Environmental Stress*. (Springer Science & Business Media, 2013).
 27. Asparouhov, T. & Muthén, B. Exploratory structural equation modeling. *Struct. Equ. Model. Multidiscip. J.* **16**, 397–438 (2009).
 28. Nurius, P. S., Uehara, E. & Zatzick, D. F. Intersection of stress, social disadvantage, and life course processes: Reframing trauma and mental health. *Am. J. Psychiatr. Rehabil.* **16**, 91–114 (2013).
 29. Matthews, K. A. & Gallo, L. C. Psychological perspectives on pathways linking socioeconomic status and physical health. *Annu. Rev. Psychol.* **62**, 501–530 (2011).
 30. Seabrook, J. A. & Avison, W. R. Socioeconomic status and cumulative disadvantage processes across the life course: Implications for health outcomes. *Can. Rev. Socio Can. Socio* **49**, 50–68 (2012).
 31. Olthof, M., Lichtwarck-Aschoff, A. & Fried, E. I. Reification of the p factor draws attention away from external causes of psychopathology. *J. Psychopathol. Clin. Sci.* <https://psycnet.apa.org/record/2025-41510-001> (2024).
 32. Maung, H. H. Externalist argument against medical assistance in dying for psychiatric illness. *J. Med. Ethics* **49**, 553–557 (2023).
 33. Buckner, R. L. The cerebellum and cognitive function: 25 years of insight from anatomy and neuroimaging. *Neuron* **80**, 807–815 (2013).
 34. Foti, F. et al. Exposure to an enriched environment accelerates recovery from cerebellar lesion. *Cerebellum* **10**, 104–119 (2011).
 35. Xu, J. et al. Effects of urban living environments on mental health in adults. *Nat. Med.* **29**, 1456–1467 (2023).
 36. Bennett, I. J. & Madden, D. J. Disconnected aging: Cerebral white matter integrity and age-related differences in cognition. *Neuroscience* **276**, 187–205 (2014).
 37. Beheshti, I., Maikusa, N. & Matsuda, H. Effects of aging on brain volumes in healthy individuals across adulthood. *Neurol. Sci.* **40**, 1191–1198 (2019).
 38. for the CHARTER Group et al. White matter damage, neuroinflammation, and neuronal integrity in HAND. *J. Neurovirol.* **25**, 32–41 (2019).
 39. Pirko, I., Lucchinetti, C. F., Sriram, S. & Bakshi, R. Gray matter involvement in multiple sclerosis. *Neurology* **68**, 634–642 (2007).
 40. Serrano-Pozo, A., Frosch, M. P., Masliah, E. & Hyman, B. T. Neuro-pathological alterations in Alzheimer disease. *Cold Spring Harb. Perspect. Med.* **1**, a006189 (2011).
 41. Wang, Y. et al. Increased risk of chronic diseases and multimorbidity in middle-aged and elderly individuals with early vision, hearing, or dual sensory impairments: insights from prospective cohort studies and Mendelian randomization analysis. *BMC Med.* **23**, 118 (2025).
 42. Levy, M. Z., Allsopp, R. C., Fletcher, A. B., Greider, C. W. & Harley, C. B. Telomere end-replication problem and cell aging. *J. Mol. Biol.* **225**, 951–960 (1992).
 43. Rahman, M. & Fukui, T. Bidi smoking and health. *Public Health* **114**, 123–127 (2000).
 44. Akl, E. A. et al. The effects of waterpipe tobacco smoking on health outcomes: a systematic review. *Int. J. Epidemiol.* **39**, 834–857 (2010).
 45. Nuttall, F. Q. Body mass index: obesity, BMI, and health: a critical review. *Nutr. Today* **50**, 117–128 (2015).
 46. Jike, M., Itani, O., Watanabe, N., Buysse, D. J. & Kaneita, Y. Long sleep duration and health outcomes: A systematic review, meta-analysis and meta-regression. *Sleep. Med. Rev.* **39**, 25–36 (2018).
 47. Sly, P. D. et al. Health consequences of environmental exposures: causal thinking in global environmental epidemiology. *Ann. Glob. Health* **82**, 3–9 (2016).
 48. Hackman, D. A., Farah, M. J. & Meaney, M. J. Socioeconomic status and the brain: mechanistic insights from human and animal research. *Nat. Rev. Neurosci.* **11**, 651–659 (2010).
 49. Xie, W. et al. Growth faltering is associated with altered brain functional connectivity and cognitive outcomes in urban Bangladeshi children exposed to early adversity. *BMC Med.* **17**, 199 (2019).
 50. Fry, A. et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. <https://doi.org/10.1093/aje/kwx246>.
 51. Cohen, P. & Cohen, J. The clinician’s illusion. *Arch. Gen. Psychiatry* **41**, 1178–1182 (1984).
 52. Sudlow, C. et al. UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLOS Med.* **12**, e1001779 (2015).
 53. Quan, H. et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med. Care* **43**, 1130–1139 (2005).
 54. Evans, A. C., Janke, A. L., Collins, D. L. & Baillet, S. Brain templates and atlases. *NeuroImage* **62**, 911–922 (2012).
 55. Diedrichsen, J., Balsters, J. H., Flavell, J., Cussans, E. & Ramnani, N. A probabilistic MR atlas of the human cerebellum. *NeuroImage* **46**, 39–46 (2009).
 56. Oishi, K. et al. Human brain white matter atlas: Identification and assignment of common anatomical structures in superficial white matter. *NeuroImage* **43**, 447–457 (2008).

57. Jordan, H., Roderick, P. & Martin, D. The index of multiple deprivation 2000 and accessibility effects on health. *J. Epidemiol. Community Health* **58**, 250–257 (2004).
58. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
59. Suhr, D. D. Exploratory or confirmatory factor analysis? <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=915520bdfde1423b0b73f7ed560c68e81678cda6> (2006).
60. Morin, A. J. S., Arens, A. K. & Marsh, H. W. A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Struct. Equ. Model. Multidiscip. J.* **23**, 116–139 (2016).
61. Stucky, B. D. & Edelen, M. O. Using hierarchical IRT models to create unidimensional measures from multidimensional data. *Handb. Item Response Theory Model.* 183–206 (2014).
62. Sijtsma, K. On the use, the misuse, and the very limited usefulness of Cronbach's alpha. <https://psycnet.apa.org/record/2016-48704-014> (2016).
63. Revelle, W. & Zinbarg, R. E. Coefficients Alpha, Beta, Omega, and the glb: Comments on Sijtsma. *Psychometrika* **74**, 145–154 (2009).
64. Reise, S. P., Bonifay, W. E. & Haviland, M. G. Scoring and modeling psychological measures in the presence of multidimensionality. *J. Pers. Assess.* **95**, 129–140 (2013).
65. Zinbarg, R. E., Revelle, W., Yovel, I. & Li, W. Cronbach's α , Revelle's β , and McDonald's ω H: their relations with each other and two alternative conceptualizations of reliability. *Psychometrika* **70**, 123–133 (2005).
66. Grice, J. W. Computing and evaluating factor scores. *Psychol. Methods* **6**, 430 (2001).
67. Hancock, G. R. & Mueller, R. O. Rethinking construct reliability within latent variable systems. *Struct. Equ. Model. Present Future* **195**, 60–70 (2001).
68. Rodriguez, A., Reise, S. P. & Haviland, M. G. Applying bifactor statistical indices in the evaluation of psychological measures. *J. Pers. Assess.* **98**, 223–237 (2016).
69. Gorsuch, R. L. *Factor Analysis: Classic Edition*. (Routledge, 2014).
70. Marquand, A. F. et al. Conceptualizing mental disorders as deviations from normative functioning. *Mol. Psychiatry* **24**, 1415–1424 (2019).
71. Wolfers, T. et al. Mapping the heterogeneous phenotype of schizophrenia and bipolar disorder using normative models. *JAMA Psychiatry* **75**, 1146–1155 (2018).
72. Marquand, A. F., Rezek, I., Buitelaar, J. & Beckmann, C. F. Understanding heterogeneity in clinical cohorts using normative models: Beyond case-control studies. *Biol. Psychiatry* **80**, 552–561 (2016).
73. Segal, A. et al. Regional, circuit and network heterogeneity of brain abnormalities in psychiatric disorders. *Nat. Neurosci.* **26**, 1613–1629 (2023).
74. Lv, J. et al. Individual deviations from normative models of brain structure in a large cross-sectional schizophrenia cohort. *Mol. Psychiatry* **26**, 3512–3523 (2021).
75. Wolfers, T. et al. Individual differences v. the average patient: mapping the heterogeneity in ADHD using normative models. *Psychol. Med.* **50**, 314–323 (2020).
76. Frazza, C. J., Dinga, R., Beckmann, C. F. & Marquand, A. F. Warped Bayesian linear regression for normative modelling of big data. *NeuroImage* **245**, 118715 (2021).
77. Lindstrom, M. J. & Bates, D. M. Newton–Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *J. Am. Stat. Assoc.* **83**, 1014–1022 (1988).
78. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
79. Haseman, J. K. & Elston, R. C. The investigation of linkage between a quantitative trait and a marker locus. *Behav. Genet.* **2**, 3–19 (1972).
80. Jiang, L. et al. A resource-efficient tool for mixed model association analysis of large-scale data. *Nat. Genet.* **51**, 1749–1755 (2019).
81. Yang, J. et al. Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19**, 807–812 (2011).
82. Watanabe, K., Taskesen, E., Van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
83. Welter, D. et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* **42**, D1001–D1006 (2014).
84. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization | Nucleic Acids Research | Oxford Academic. https://academic.oup.com/nar/article/37/suppl_2/W305/1149611.
85. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
86. Anderson, B. O. et al. Health and cancer risks associated with low levels of alcohol consumption. *Lancet Public Health* **8**, e6–e7 (2023).
87. Prince, M. et al. No health without mental health. *lancet* **370**, 859–877 (2007).
88. Jiang, L., Zheng, Z., Fang, H. & Yang, J. A generalized linear mixed model association tool for biobank-scale data. *Nat. Genet.* **53**, 1616–1621 (2021).
89. Hayes, A. F. *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. (Guilford publications, 2017).
90. Chen, J. Code for 'Neural-genetic-environmental evidence for a disease factor in mental and physical health multimorbidity'. <https://doi.org/10.5281/zenodo.17280644> (2025).

Acknowledgements

This work was supported by the Yunnan Fundamental Research Projects (Grant Nos. 202501AV070005 to J. Wang, 202201BE070001-004 to J. Wang) and the National Natural Science Foundation of China (Grant No. 62176044 to J. Wang).

Author contributions

J. W., S. C., and C. C. contributed to the conception and design of the study. J. C., C. C., and W. L. contributed to data processing and analyzing. J. C. and M. G.-A. contributed to structural equation modeling. S. L. contributed to GWAS analyzing. J. C., C. C., J. W., and Y. Z. wrote and edited the manuscript. C. C., J. W., S. C., M. S., T. J., J. K., H. S., and W. L. provided significant feedback on the analyses and the manuscript. J. Chen produced the figures. All the authors commented and discussed the results.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-65055-w>.

Correspondence and requests for materials should be addressed to Congying Chu, Samuele Cortese or Jiaojian Wang.

Peer review information *Nature Communications* thanks Ruoyu Tian and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025