A comprehensive meta-analysis of human assortative mating in 22 complex traits

Tanya B Horwitz[1,2*] and Matthew C Keller[1,2*]

[1]Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, CO, United States of America.

[2]Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, CO, United States of America.

Abstract

Assortative mating (AM) occurs when the correlation for a trait between mates is larger than would be expected by chance. AM can increase the genetic and environmental variation of traits, can increase the prevalence of disorders in a population, and can bias estimates in genetically informed designs. In this study, we conducted the largest set of meta-analyses on human AM published to date. Across 22 traits, meta-analyzed correlations ranged from $r = .08$ to $r = .58$, with social attitude, substance use, and cognitive traits showing the highest correlations and personality, disorder, and biometrical traits generally yielding smaller but still positive and nominally significant ($p < .05$) correlations. We observed high between-study heterogeneity for most traits, which could have been the result of phenotypic measurement differences between samples and/or differences in the degree of AM across time or cultures.

A comprehensive meta-analysis of human assortative mating in 22 complex traits

Assortative mating (AM) is the phenomenon whereby individuals with similar trait values mate with one another at levels higher than expected by chance[1]. Contrary to the maxim "opposites attract," nonzero phenotypic correlations between human[2–21] and nonhuman[1] mates are overwhelmingly in the positive direction, with only a handful of examples of disassortative mating, or negative mate correlations, reported in the literature[1,4,8,20,22–29]. Several potential mechanisms of AM in humans have been described, although they are not mutually exclusive because multiple mechanisms can simultaneously be responsible for observed correlations. Phenotypic homogamy (also known as primary phenotypic assortment) occurs when mates match directly on the trait of interest[30]. While phenotypic homogamy is often conceptualized as mates actively preferring similarity, this type of homogamy can also be a function of indirect selection, such as when mates are chosen from among strata that are partially determined by individuals' phenotypic values (e.g., AM for educational attainment arising as an indirect consequence of mate choice occurring within job occupations). Social homogamy, on the other hand, occurs when individuals match within strata that are determined by non-heritable background social factors[18,31], such as within social class in cultures where class is not genetically influenced. At the other end of the spectrum, genetic homogamy is the mechanism whereby mates correlate more genetically than phenotypically for a trait; this can occur when there is phenotypic homogamy on a trait that is more correlated genetically than environmentally with the trait of interest[30,32]. Finally, convergence occurs when mates become more similar over time[3,8], either due to direct (reciprocal or one-way) phenotypic influences on one another or to the mutual influence of shared environmental factors.

Social scientists and quantitative geneticists care about the mechanisms and the strength of AM because both influence parameters of interest and impact how various estimates in the literature should be interpreted. Phenotypic and genetic homogamy on heritable traits increase correlations between and within causal loci, which in turn increases the genetic covariance between relatives and the trait's phenotypic and genetic variation. Such an increase in variation could manifest as increased prevalence rates of dichotomous traits such as psychiatric disorders[18,33], although this effect should only be pronounced in rare, highly heritable disorders under strong AM[18]. Social homogamy can also increase trait variation when parental phenotypic values for sociocultural traits are inherited by offspring via vertical transmission[34]. Failing to account for AM can lead to biases in estimates from genetically informed designs, including the association statistics from genome-wide association studies[35], heritability estimates from twin/family designs and from single nucleotide polymorphisms[36], and the strength of estimated causal associations in Mendelian randomization studies[37].

Given that the genetic consequences of AM and the impacts of not accounting for it in certain genetically informed designs are non-negligible, it is important to understand the strength of AM for traits commonly investigated in human genetics. The strength and breadth of AM is also of interest to investigators of human mating in psychology, sociology, and economics. While many studies have reported estimates of AM in humans, we are aware of no study that has meta-analyzed AM on a large number of phenotypically diverse traits. In the current report, we use stringent methodology to meta-analyze and compare partner correlations for 22 commonly investigated complex traits. These results are the most comprehensive set of meta-analyses on human AM to date, and should shed light on contemporary human mating trends, help with the

interpretation of heritability estimates, motivate studies into the various causes of AM across traits, and aid in the choice of design in genetic studies.

Results

Meta-analysis

We meta-analyzed partner concordance rates for 22 traits. While AM has been analyzed for hundreds of traits, we focused on those most studied in the AM literature as well as some less commonly studied dichotomous traits that have important health implications. The total number of partner pairs for each trait ranged from 2,270 (for drinking quantity) to 1,533,956 (for substance use disorder); effective sample sizes for dichotomous traits (see *Methods*) ranged from 721 (for alcohol use disorder) to 241,817 (for substance use disorder). Supplementary Tables S1 and S2 show all studies that we included in our meta-analysis for continuous and dichotomous traits, respectively, as well as the effect sizes for each sample. For comparability across traits, we focus here on Pearson and tetrachoric correlations for continuous and dichotomous traits, respectively. Supplementary Table S2 also includes an alternative metric of partner concordance for dichotomous traits, the odds ratio (OR), which is the odds of a participant possessing a trait given that their partner has it divided by the odds of a participant possessing the trait given that their partner does not have it. Supplementary Table S3 lists studies excluded from our meta-analysis along with the reasons for their exclusion.

Fig. 1 displays the meta-analyzed random effects correlations for all traits along with their 95% confidence intervals. The meta-analyzed correlations were greater than zero at the nominal significance level ($p < .05$) for all traits. The point estimates for fourteen traits were also

significant at the Bonferroni-corrected ($p < .05/22 = 0.00227$) significance level. Cognitive and social attitude traits showed the highest correlations ($.39 \leq r_{meta} \leq .58$); personality, anthropometric traits, substance use disorders, and other disorders showed the lowest ($.08 \leq r_{meta} \leq .29$); and correlations for non-pathological substance use traits typically lay between these two sets ($.24 \leq r_{meta} \leq .54$) (see Table 1). Fig. S1 displays forest plots for all the traits we analyzed with publications ordered by year and color-coded by region. The meta-analyzed fixed effects results for each trait (Fig. S2) were qualitatively similar to the random effect results. Fig. S5 shows the number of studies included and excluded for each trait.

Table 2 summarizes each trait's heterogeneity estimates and the prediction intervals of future studies' effect sizes. We quantified heterogeneity using the Higgins & Thompson's I^2 metric, which represents the percentage of variance resulting from between-study heterogeneity in effect sizes rather than within-study sampling error[38]. Higgins and Thompson (2002)[39] classified I^2 values of 25%, 50%, and 75% as low, medium, and high heterogeneity, respectively. Across traits in our 22 meta-analyses, the median Higgins & Thompson I^2 statistic was 87.5%, reflecting very high heterogeneity in AM estimates for most traits. However, a high I^2 reflects not only high between-study heterogeneity in estimated effect sizes but also low within-study heterogeneity due to highly precise estimates of individual studies. Thus, these high I^2 values may in part be due to the high precision of estimates afforded by the large sample sizes of many of the studies included in our analyses. An alternative metric of heterogeneity that is unaffected by the precision of estimates of individual studies, τ^2, represents the estimated variance of the true effect size under a random effects model. The estimated standard deviations of true effects (τ) were large relative to the meta-analyzed correlation values for many traits. The median coefficient of variation $\left(\frac{\tau}{r_{meta}}\right)$ was .41, and the coefficient of variation was above .50 for

intelligence quotient (IQ), drinking quantity, agreeableness, conscientiousness, extraversion, body mass index (BMI), and generalized anxiety disorder (GAD). However, for some traits, such as EA (r_{meta} = .53 +/- τ = .10), political values (r_{meta} = .58 +/- .08), and depression (r_{meta} = .14 +/- .02), the estimated standard deviation of true effects was not very large compared to the meta-analyzed estimate. Overall, our results suggest that AM is characterized by substantial differences in the strength of true effect across populations differentiated by place or time.

For each trait, we also created Graphic Display of Heterogeneity (GOSH) plots (Fig. S4)[40], which are scatterplots of the meta-analyzed correlations for all possible 2^{k-1} combinations of k studies of size 2 through k (up to 1 million combinations) on the x-axis and the I^2 values of these combinations on the y-axis. Two or more distinct clusters anywhere in the plot may indicate subpopulations that differ in their average effect size[40], although a smear of points along the bottom of GOSH plots is caused by two or more study results that happen to be similar (thereby producing I^2 values near 0) and is typically not of interest. For most traits plotted in Fig. S4, there are no obvious clusters. However, for IQ and conscientiousness, there do appear to be two clusters, one made up of study combinations that have higher heterogeneity and higher average correlations, and another with lower heterogeneity and lower average correlations. The two clusters in the GOSH plot for IQ may have resulted from an outlier reported in a 1938 study that found a partner correlation of .81[41], which is substantially greater than the meta-analyzed estimate we report for this trait.

Because AM studies ostensibly focus more on effect size than hypothesis testing, we expected that publication bias was unlikely to be a major factor for the study results we meta-analyzed. Nevertheless, we created funnel plots (Fig. S3), which plot study effect size (Fisher Z transformed correlations here) on the x-axis against standard error on the y-axis, to visually

inspect whether there was evidence for asymmetry, a potential indicator of publication bias. Overall, there was no obvious asymmetry across the funnel plots. Only for IQ and drinking quantity did it appear that there may be a systematic bias of larger studies having smaller effect sizes, but both were based on 10 or fewer studies, which can lead to apparent asymmetry by chance[38,42]. The more obvious pattern observed in most funnel plots was the large number of points that were outside the expected triangular region, again reflecting the high heterogeneity in correlations observed across studies.

Discussion

In this study, we collated and synthesized the results from a large number of studies on human AM to provide a better understanding of which traits mates assort on and how strong the assortment is. To our knowledge, this is the largest and most comprehensive set of meta-analyses on human AM to date. We found the highest levels of AM for political and religious values, educational attainment, IQ, and some substance use traits; partner correlations for other traits were smaller. Nevertheless, we found nominally significant ($p < .05$) evidence for AM for every trait investigated. More than half of the meta-analyzed correlations were also significant at the Bonferroni-corrected level. Whether these correlations are due to convergence or to initial nonrandom mating based on phenotypic, social, or genetic homogamy remains to be determined, though some research has attempted to investigate which of these mechanisms is responsible for observed AM for particular traits.

The two social attitude traits that we examined—political attitudes and religiosity—showed the highest levels of AM of all the traits we assessed. For these traits, we examined continuous measures of attitudes toward political issues and self-report of multiple religious

ideas/practices. Interestingly, despite clear geographical stratification of religious and voting trends apparent in countries such as the United States, most studies to date investigating the cause of mate similarity on political and religious attitudes have suggested that the data is most consistent with phenotypic rather than social homogamy, and there is no compelling evidence of substantial convergence for either trait[4,43–46]. This may be relevant to current events because, to the degree that social attitudes are genetically or socially heritable, AM on them may contribute to heightened political and cultural polarization.

We also found a high partner correlations for educational attainment (EA) (r_{meta} = .53), and only one sample[47] out of 27 reported a correlation under .30. Thus, there is consistent evidence for strong AM on EA across recent decades and across cultures in which the trait has been studied. Robinson *et al.* (2017)[32] found that the implied phenotypic correlation for EA between partners in the UK Biobank, extrapolated from the observed correlation between partners' trait-associated loci, was .65. This value was substantially larger than the phenotypic correlation they observed for EA in the same sample and exceeds the upper limit of our confidence interval for the meta-analyzed EA partner correlation. This suggests that AM for EA is consistent with genetic homogamy, and that mates may be assorting on some trait that is more genetically than environmentally correlated with EA. Contrary to Robinson *et al.*'s (2017)[32] finding, Torvik *et al.* (2022)[48] did not find evidence for genetic homogamy in educational attainment in a sample of partners, siblings, and in-laws in Norway. Instead, they found evidence that AM on EA was due to a mix of both social homogamy and phenotypic homogamy. Whether this discrepancy is due to differences in EA AM between Norway and the UK or to differences in sample characteristics (e.g., ascertainment) is an open question.

The meta-analyzed partner correlation coefficients for substance use/abuse traits ranged from r_{meta} = .24 to r_{meta} = .54. Interestingly, some (but not all[49,50]) studies that have examined mechanisms of assortment in drinking and smoking have reported evidence of convergence for these behaviors[6,8,12,51], making these traits amongst the only ones to show support for convergence in the literature.

We observed substantial between-study heterogeneity in partner correlations for most traits. A large degree of between-study heterogeneity would certainly be problematic in fixed effects meta-analyses that assume a single underlying effect. However, even for random effects meta-analyses, which are viewed as more appropriate when heterogeneity is present, high levels of heterogeneity suggest caution should be used in interpretation of results. Random effects meta-analyses assume an underlying (normal) distribution of true effects across the studies' sampled populations, and the meta-analytic result is the estimated mean of those true effects. Thus, the estimates we present here cannot be interpreted as estimates of a single true level of AM for a given trait, but rather estimates of the typical level of AM across many possible levels that might be observed at different times or locations.

There are several possible causes of the high levels of heterogeneity in AM we observed across studies within the same trait. Most obviously, it is possible that the true degree of AM varied across populations due to cultural differences in mating systems or preferences. This seems plausible; AM involves mate preferences, social stratification, and/or couple dynamics, and so it is unlikely to be consistent across different cultural contexts. Differences in population size, mobility, and/or education across populations may impact the pool of a person's potential mates and thereby the degree to which preferences can be acted on. However, there was insufficient cultural diversity within traits to test whether there were significant differences in

partner concordance across cultures. Similarly, we determined that publication year was too coarse a metric of the year in which mates were married, and too many studies failed to report sufficient information for us to formally assess changes in AM over time.

It is also possible that some of the heterogeneity in AM effect sizes was due to differences in how constructs were measured across studies—for example, differences in the measurement batteries used, differences in participants' interpretations of battery items, or differences in the clinical thresholds employed. Potentially consistent with this possibility, we observed that the prevalence rates of dichotomous traits varied greatly in supposedly non-ascertained samples, which may have contributed to the heterogeneity we observed in our correlation coefficients. Nevertheless, we observed high levels of heterogeneity even for traits—such as height and BMI—measured in standardized ways, suggesting that differences in how the constructs were measured is unlikely to be a complete explanation. Finally, it is possible that publication bias led to heterogeneity, particularly if studies that found AM results that were substantially different from those already published in the literature were more likely to be submitted and published—a kind of "novelty bias." However, it is also possible that a "conformity bias" exists in the opposite direction and has led to downwardly biased estimates of heterogeneity. While we could not test and therefore cannot rule out either possibility, we find them unlikely given that the incentives for both seem dubious.

Although we initially gathered data on AM for rare psychiatric disorders, we did not formally meta-analyze the tetrachoric correlations for these traits because too few studies met our inclusion criteria as a result of unspecified sample sizes, the use of longitudinal rather than cross-sectional measurements of concordance, and small expected cell frequencies (see Supplementary Table S2 and S3). Nevertheless, studies that have provided robust estimates of

partner concordance for psychiatric disorders have suggested low to moderate AM, both within and across disorders[18,21,52,53]. For example, based on data from Swedish population registers that included more than 700,000 unique cases—originally analyzed by Nordsletten *et al.* (2016)[54]--Peyrot *et al.* (2016)[18] estimated ascertainment-corrected tetrachoric correlation coefficients of .26 for schizophrenia, .10 for bipolar disorder, .28 for autism spectrum disorder, and .31 for attention-deficit/hyperactivity disorder.

There are several implications for the consistent evidence of AM across traits we documented in this meta-analysis. First, as noted above, AM can increase the genetic variance and the prevalence of a disorder. Although the increase in prevalence for common disorders may not be large (e.g., ~10%), the levels of AM observed for rare traits of high heritability, such as autism, could lead to a ~1.5-fold prevalence increase after one generation, and an even higher increase (~2.4-fold) over many generations[18]. Second, AM can create biases in estimates of interest in genetically informative designs, such as estimates based on twin studies[10,54], genome-wide association studies (GWAS)[35], Mendelian randomization[37], and SNP-heritability[36]. Finally, to the degree that the heterogeneity in AM we observed was due to true differences in the strength of AM rather than differences in measurement, our estimates of the strength of AM may not generalize to other populations. While estimates for some traits, such as height, were based on a geographically and ethnically diverse set of samples, most of the samples included in our meta-analyses were drawn from Europe, North America, and Australia, and Asia. For example, all estimates of AM for religiosity came from samples in the United States.

In summary, we conducted the largest and most comprehensive set of meta-analyses of human AM to date. Our estimates were based on nearly a century of research and millions of partner pairs. We found high partner correlations for traits related to substance use, IQ, EA, and

social attitudes, and smaller but nominally significant ($p < .05$) correlations for personality, anthropometric, and disorder traits. However, we also observed high levels of heterogeneity in AM estimates across studies for most traits investigated, suggesting that AM may differ across time or place and that a single estimate of AM cannot typically be assumed for a given trait across populations.

Methods

Inclusion and exclusion criteria

We conducted a systematic review of English-language studies that examined AM based on partners' continuous and dichotomous self-reports on the same complex traits. All included studies were published in peer-reviewed journals on or before December 22, 2021. To conduct this review, we searched for words pertaining to the traits of interest in conjunction with the terms *assortative mating, assortative marriage, partner concordance, partner correlation, nonrandom mating, homogamy, marital resemblance,* and *marital homophily* in Google Scholar, and we checked relevant papers cited in these studies for adherence to our criteria. We restricted our analysis to studies of opposite-sex co-parents, engaged pairs, married pairs, and/or cohabitating pairs (referred to as "partners" hereafter), with a few studies containing a small number of divorced couples; we excluded same-sex partners because same-sex and opposite-sex pairs show different patterns of assortment for some traits[55,56], because there is less data on the former, and because same-sex assortment does not have the same implications for genetic studies. With the exception of studies that intentionally ascertained partners for the trait of interest, we excluded studies in which pairs had a characteristic that deviated from the norm in the general population in a way that might have affected the magnitude of concordance (e.g., a

sample of only adoptive parents was excluded), and we only included studies where the sample size was reported or could be inferred. For example, if only percentages were reported for each cell of a contingency table, the sample size of each cell could be inferred as the percentage multiplied by *N*.

We restricted our analysis to studies with sample sizes greater than 100. For dichotomous traits, we restricted our analysis to studies with expected contingency table cell frequencies of five or greater and observed cell frequencies greater than zero. When the samples in multiple studies that were appropriate for our meta-analysis overlapped or were likely to have overlapped based on information provided in the publication, we only used the study with the largest sample size. We calculated effect sizes from the data reported in primary studies rather than relying on effect size estimates from other published meta-analyses. If a study reported partner concordance rates for multiple independent samples, each was included as a separate entry. When studies reported partner correlation at different waves, we reported the results from the first wave.

When studies reported both the raw correlation and the partial correlation(s) controlling for covariates (such as age), we included the raw correlation for consistency across studies. For studies that only reported partial correlations, we used the estimate with the fewest number of covariates. For ordinal and continuous traits, studies typically reported Spearman's rho or Pearson's *r* but at times reported polychoric correlations. We excluded polychoric correlations reported for such traits in order to avoid pooling two classes of correlation for the same meta-analyzed effect size. Because polychoric correlations occurred rarely, we do not anticipate a large loss of power as a result. Because AM for height has already been meta-analyzed extensively by Stulp *et al.* (2017)[9], we re-analyzed studies from the paper's supplement in the

same way we analyzed other continuous traits, after eliminating studies from this meta-analysis in accordance with our exclusion criteria. Finally, we restricted our meta-analysis to traits for which there were at least three samples that met our criteria.

Dichotomous traits

For dichotomous traits, we primarily considered studies that examined pairs in non-ascertained community samples or national registers as well as those from samples that ascertained probands. Most ascertained studies were ultimately excluded because probands were typically in clinical settings (e.g., hospitalized), whereas partners of probands with the disorder typically were not. Although such ascertainment can be dealt with if all the applicable populations' (i.e. inpatient, outpatient, and those who have never received treatment) prevalence rates are known, it was typically impossible to know all of these rates. We eliminated any ascertained studies in which there was a >~two-fold difference in male and female prevalence if there was not enough information to divide discordant couples based on sex. Simulation results suggested that mixing individuals of different sexes when prevalence rates were more discrepant than this would lead to unacceptable levels of bias. Because of possible differences in the strength of AM implied from concordance of male probands versus that implied from female probands, we excluded studies that only included single-sex probands. When both male and female proband data was available (only a single study[52]), estimates based on each proband (female and male) were included as separate results.

We only used cross-sectional measures of partner concordance and therefore excluded studies that used longitudinal metrics such as morbidity risks[57], hazard ratios, and incidence ratios. We required that either odds ratios (ORs), risk ratios (RR), phi coefficients (Φ), contingency tables, or—if the study was not ascertained (see below)–tetrachoric correlations,

were reported for dichotomous traits. Concordance rates captured by any of the first four of these measures were then converted to tetrachoric correlations for consistency. When the contingency table was unknown but the OR was reported, we first inferred the contingency table using an R function described in the supplementary methods of Peyrot *et al.* (2016)[18] (provided to us by the authors) and then estimated the tetrachoric correlation. When the contingency table was provided, we calculated the OR and tetrachoric correlation (using the polychoric() function from the "polycor" package[58]) in R ourselves, and thus the effect size we used in our analysis was sometimes different than that reported in the original study. When the contingency table was unknown but Φ was reported, Φ was converted to a tetrachoric correlation using the phi2tetra() function from the "psych" package[59] in R. The prevalence rates for each sex used for these conversions (from Φ and the OR) are reported in Supplementary Table S2. No studies that we included in our final analysis reported an RR.

For studies where probands were ascertained, we used the OR, which is not influenced by ascertainment, along with estimates of sex-specific prevalence rates from the country or region the sample came from, to calculate tetrachoric correlations. To do this, we used the aforementioned R function provided to us by Peyrot and colleagues, which produces the population (non-ascertained) contingency table that is implied given the observed OR in the ascertained sample and the assumed population prevalence in each sex. We then used this implied contingency table to estimate the underlying (non-ascertained) tetrachoric correlation in the population. This correction is necessary because the liability in the ascertained sample, where the case to control ratio is usually higher than that in the population, is different than the liability distribution in the population, which would lead to upwardly biased estimates if the tetrachoric correlation was estimated based on just the sample contingency table.

We used the metacor() function from the "meta" package in R[60] to conduct both random and fixed effects meta-analyses using inverse-variance weighting of the Fisher z transformed correlations. For continuous traits, we used the Knapp-Hartung adjustment[61,62] to calculate the variance of point estimates and restricted maximum-likelihood (REML) to estimate τ^2, the variance of the true overall effect size under random effects[63,64]. For binary traits, we used the Paule-Mandel estimator[65] to estimate τ^2 and applied the Knapp-Hartung adjustment[61,62] to our calculation of the variance of the point estimate. We conducted a Monte Carlo analysis to determine how best to pool information for different studies in a meta-analysis. While the "true" base spousal correlation varied across simulated meta-analyses, the population-level spousal correlation across "studies" within the same meta-analysis was consistent (in order to establish a true rate of spousal concordance against which to compare our point estimates). However, prevalence rates were allowed to vary across populations in the same simulated meta-analysis (see Supplementary Table S4 for the results of each method used in conjunction with various parameter estimates). We found that calculating tetrachoric correlations for each sample and then meta-analyzing them provided more accurate point estimates than pooling contingency tables and then calculating tetrachoric correlations. Thus, we followed this procedure for binary traits throughout. The metacor() function internally calculates the expected variance of correlations based on sample sizes and assumes they are Pearson correlations, which would be incorrect for tetrachoric correlations. Thus, we needed to input effective (rather than actual) sample sizes for tetrachoric correlations. For non-ascertained studies, we estimated the effective sample sizes by using the standard error calculated in the polychor() package and solving for *n* in the equation $SE(r) = \sqrt{\frac{(1-r^2)}{(n-2)}}$. For ascertained studies examining dichotomous traits, we created bootstrapped contingency tables, each of size *n* (the number of partners) and sampled from the study's (raw,

ascertained) contingency table with replacement. We followed the procedure described above to convert the ascertained contingency table to a tetrachoric correlation corrected for ascertainment. We repeated this process 1,000 times, calculated the standard error by estimating the standard deviation of the 1,000 bootstrapped tetrachoric correlations, and used this standard error to calculate the effective sample size as described above.

Four of the traits in our supplementary tables—bipolar disorder, schizophrenia, panic disorder, and phobia—posed a problem because they were rare (bipolar disorder and schizophrenia) or have not been studied in sufficiently large samples (panic disorder and phobia). This resulted in contingency tables with zero frequency cells or with expected cell frequencies that were less than five. As a result, there was not a sufficient number of studies meeting our inclusion criteria to justify formally meta-analyzing these four traits, though we included the results from studies that otherwise met our criteria for these traits in Supplementary Table S2.

Data availability

Studies included in the meta-analysis are listed in Supplementary Tables S1 and S2, and studies excluded from the meta-analysis are listed in Supplementary Table S3.

Code availability

The code for the analyses and simulations is available from the authors upon request.

References

1. Jiang, Y., Bolnick, D. I. & Kirkpatrick, M. Assortative mating in animals. *Am. Nat.* **181**, E125–E138 (2013).
2. Luo, S. Assortative mating and couple similarity: patterns, mechanisms, and consequences. *Soc. Personal. Psychol. Compass* **11**, e12337 (2017).
3. Burgess, E. W. & Wallin, P. Homogamy in social characteristics. *Am. J. Sociol.* **49**, 109–124 (1943).
4. Alford, J. R., Hatemi, P. K., Hibbing, J. R., Martin, N. G. & Eaves, L. J. The politics of mate choice. *J. Polit.* **73**, 362–379 (2011).
5. Jurj, A. L. *et al.* Spousal correlations for lifestyle factors and selected diseases in Chinese couples. *Ann. Epidemiol.* **16**, 285–291 (2006).
6. Ask, H., Rognmo, K., Torvik, F. A., Røysamb, E. & Tambs, K. Non-random mating and convergence over time for alcohol consumption, smoking, and exercise: the Nord-Trøndelag Health Study. *Behav. Genet.* **42**, 354–365 (2012).
7. Rawlik, K., Canela-Xandri, O. & Tenesa, A. Indirect assortative mating for human disease and longevity. *Heredity* **123**, 106–116 (2019).
8. Price, R. A. & Vandenberg, S. G. Spouse similarity in American and Swedish couples. *Behav. Genet.* **10**, 59–71 (1980).
9. Stulp, G., Simons, M. J. P., Grasman, S. & Pollet, T. V. Assortative mating for human height: a meta-analysis. *Am. J. Hum. Biol.* **29**, e22917 (2017).
10. Allison, D. B. *et al.* Assortative mating for relative weight: genetic implications. *Behav. Genet.* **26**, 103–111 (1996).

412 11. Mascie-Taylor, C. G. N. Spouse similarity for IQ and personality and convergence. *Behav.*
413 *Genet.* **19**, 223–227 (1989).
414 12. Meyler, D., Stimpson, J. P. & Peek, M. K. Health concordance within couples: a systematic
415 review. *Soc. Sci. Med.* **64**, 2297–2310 (2007).
416 13. Stimpson, J. P. & Peek, M. K. Concordance of chronic conditions in older Mexican
417 American couples. *Prev. Chronic. Dis.* **2**, (2005).
418 14. Jeong, S. & Cho, S.-I. Concordance in the health behaviors of couples by age: a cross-
419 sectional study. *J. Prev. Med. Pub. Health* **51**, 6 (2018).
420 15. Di Castelnuovo, A., Quacquaruccio, G., Donati, M. B., de Gaetano, G. & Iacoviello, L.
421 Spousal concordance for major coronary risk factors: a systematic review and meta-analysis.
422 *Am. J. Epidemiol.* **169**, 1–8 (2009).
423 16. Hippisley-Cox, J. Married couples' risk of same disease: cross sectional study. *BMJ* **325**,
424 636–636 (2002).
425 17. Galbaud Du Fort, G., Bland, R. C., Newman, S. C. & Boothroyd, L. J. Spouse similarity for
426 lifetime psychiatric history in the general population. *Psychol. Med.* **28**, 789–802 (1998).
427 18. Peyrot, W. J., Robinson, M. R., Penninx, B. W. J. H. & Wray, N. R. Exploring boundaries
428 for the genetic consequences of assortative mating for psychiatric traits. *JAMA Psychiatry*
429 **73**, 1189–1195 (2016).
430 19. McLeod, J. D. Social and psychological bases of homogamy for common psychiatric
431 disorders. *J. Marriage Fam.* 201–214 (1995).
432 20. Maes, H. H. *et al.* Assortative mating for major psychiatric diagnoses in two population-
433 based samples. *Psychol. Med.* **28**, 1389–1401 (1998).

21. Mathews, C. A. & Reus, V. I. Assortative mating in the affective disorders: a systematic review and meta-analysis. *Compr. Psychiatry* **42**, 257–262 (2001).

22. Watson, D. *et al.* Match makers and deal breakers: analyses of assortative mating in newlywed couples. *J. Pers.* **72**, 1029–1068 (2004).

23. Botwin, M. D., Buss, D. M. & Shackelford, T. K. Personality and mate preferences: five factors in mate selection and marital satisfaction. *J. Pers.* **65**, 107–136 (1997).

24. Eysenck, H. J. & Wakefield, J. A. Psychological factors as predictors of marital satisfaction. *Adv. Behav. Res. Ther.* **3**, 151–192 (1981).

25. McCrae, R. R. *et al.* Personality trait similarity between spouses in four cultures. *J. Pers.* **76**, 1137–1164 (2008).

26. Markey, P. M. & Markey, C. N. Romantic ideals, romantic obtainment, and relationship experiences: the complementarity of interpersonal traits among romantic partners. *J. Soc. Pers. Relatsh.* **24**, 517–533 (2007).

27. Stimpson, J. P., Masel, M. C., Rudkin, L. & Peek, M. K. Shared health behaviors among older Mexican American spouses. *Am. J. Health Behav.* **30**, 495–502 (2006).

28. Al-Sharbatti, S. S., Abed, Y. I., Al-Heety, L. M. & Basha, S. A. Spousal concordance of diabetes mellitus among women in Ajman, United Arab Emirates. *Sultan Qaboos Univ. Med. J.* **16**, e197 (2016).

29. Ober, C. *et al.* HLA and mate choice in humans. *Am. J. Hum. Genet.* **61**, 497–504 (1997).

30. Keller, M. C., Medland, S. E. & Duncan, L. E. Are extended twin family designs worth the trouble? A comparison of the bias, precision, and accuracy of parameters estimated in four twin family models. *Behav. Genet.* **40**, 377–393 (2010).

31. Cloninger, C. R. Interpretation of intrinsic and extrinsic structural relations by path analysis: theory and applications to assortative mating. *Genet. Res.* **36**, 133–145 (1980).
32. Robinson, M. R. *et al.* Genetic evidence of assortative mating in humans. *Nat. Hum. Behav.* **1**, 1–13 (2017).
33. Yengo, L. *et al.* Imprint of assortative mating on the human genome. *Nat. Hum. Behav.* **2**, 948–954 (2018).
34. Cloninger, C. R., Rice, J. & Reich, T. Multifactorial inheritance with cultural transmission and assortative mating. II. a general model of combined polygenic and cultural inheritance. *Am. J. Hum. Genet.* **31**, 176–198 (1979).
35. Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).
36. Border, R. *et al.* Assortative mating biases marker-based heritability estimators. *Nat. Commun.* **13**, 660 (2022).
37. Hartwig, F. P., Davies, N. M. & Davey Smith, G. Bias in Mendelian randomization due to assortative mating. *Genet. Epidemiol.* **42**, 608–620 (2018).
38. Harrer, M., Cuijpers, P., Furukawa, T. A. & Ebert, D. D. *Doing Meta-Analysis in R*. (2019).
39. Higgins, J. P. & Thompson, S. G. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* **21**, 1539–1558 (2002).
40. Olkin, I., Dahabreh, I. J. & Trikalinos, T. A. GOSH – a graphical display of study heterogeneity. *Res. Synth. Methods* **3**, 214–223 (2012).
41. Cattell, R. B. & Willson, J. L. Contributions concerning mental inheritance: I. of intelligence. *Br. J. Educ. Psychol.* **8**, 129–149 (1938).

42. Sterne, J. A. C. *et al.* Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* **343**, d4002 (2011).

43. Watson, D. *et al.* Match Makers and Deal Breakers: Analyses of Assortative Mating in Newlywed Couples. *J. Pers.* **72**, 1029–1068 (2004).

44. Feng, D. & Baker, L. Spouse similarity in attitudes, personality, and psychological well-being. *Behav. Genet.* **24**, 357–364 (1994).

45. Zietsch, B. P., Verweij, K. J., Heath, A. C. & Martin, N. G. Variation in human mate choice: simultaneously investigating heritability, parental influence, sexual imprinting, and assortative mating. *Am. Nat.* **177**, 605–616 (2011).

46. Martin, N. G. *et al.* Transmission of social attitudes. *Proc. Natl. Acad. Sci.* **83**, 4364–4368 (1986).

47. Procidano, M. E. & Rogler, L. H. Homogamous assortative mating among Puerto Rican families: intergenerational processes and the migration experience. *Behav. Genet.* **19**, 343–354 (1989).

48. Torvik, F. A. *et al.* Modeling assortative mating and genetic similarities between partners, siblings, and in-laws. *Nat. Commun.* **13**, 1108 (2022).

49. Treur, J. L., Vink, J. M., Boomsma, D. I. & Middeldorp, C. M. Spousal resemblance for smoking: underlying mechanisms and effects of cohort and age. *Drug Alcohol Depend.* **153**, 221–228 (2015).

50. Agrawal, A. *et al.* Assortative mating for cigarette smoking and for alcohol consumption in female Australian twins and their spouses. *Behav. Genet.* **36**, 553–566 (2006).

51. Reczek, C., Pudrovska, T., Carr, D., Thomeer, M. B. & Umberson, D. Marital histories and heavy alcohol use among older adults. *J. Health Soc. Behav.* **57**, 77–96 (2016).

52. Nordsletten, A. E. *et al.* Patterns of nonrandom mating within and across 11 major psychiatric disorders. *JAMA Psychiatry* **73**, 354–361 (2016).

53. Chou, I.-J. *et al.* Familial aggregation and heritability of schizophrenia and co-aggregation of psychiatric illnesses in affected families. *Schizophr. Bull.* **43**, 1070–1078 (2017).

54. Coventry, W. L. & Keller, M. C. Estimating the extent of parameter bias in the classical twin design: a comparison of parameter estimates from extended twin-family and classical twin designs. *Twin Res. Hum. Genet.* **8**, 214–223 (2005).

55. Schwartz, C. R. & Graf, N. L. Assortative matching among same-sex and different-sex couples in the United States, 1990–2000. *Demogr. Res.* **21**, 843 (2009).

56. Verbakel, E. & Kalmijn, M. Assortative mating among Dutch married and cohabiting same-sex and different-sex couples. *J. Marriage Fam.* **76**, 1–12 (2014).

57. Strömgren, E. Zum Ersatz des Weinbergschen "abgekürzten Verfahrens ". *Z. Für Gesamte Neurol. Psychiatr.* **153**, 784–797 (1935).

58. polycor: Polychoric and Polyserial Correlations version 0.8-1 from R-Forge. https://rdrr.io/rforge/polycor/.

59. Revelle, W. *psych: Procedures for Psychological, Psychometric, and Personality Research*. (2021).

60. Balduzzi, S., Rücker, G. & Schwarzer, G. How to perform a meta-analysis with R: a practical tutorial. *Evid. Based Ment. Health* **22**, 153–160 (2019).

61. Knapp, G. & Hartung, J. Improved tests for a random effects meta-regression with a single covariate. *Stat. Med.* **22**, 2693–2710 (2003).

62. Hartung, J. & Knapp, G. On tests of the overall treatment effect in meta-analysis with normally distributed responses. *Stat. Med.* **20**, 1771–1782 (2001).

63. Harrer, M., Cuijpers, P., Furukawa, T. A. & Ebert, D. D. *Doing Meta-Analysis in R*. (2019).

64. Veroniki, A. A. *et al.* Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Res. Synth. Methods* **7**, 55–79 (2016).

65. Paule, R. C. & Mandel, J. Consensus values and weighting factors. *J. Res. Natl. Bur. Stand.* **87**, 377–385 (1982).

Author contributions statement

TBH contributed to study design, statistical analyses, manuscript writing, collection of studies to be meta-analyzed, simulation, and creation of all figures and tables; MCK contributed to study design, statistical analyses, manuscript writing, and simulation.

Additional information

The authors declare no competing interests.

Trait	***r* [CI]**	***K***	***N***	***Effective N***	***p*-value**
EA	.53 [.49; .56]	27	230,915	NA	< .0001
IQ	.39 [.21; .54]	10	2,561	NA	.0012
Political values	.58 [.53; .63]	9	10,694	NA	< .0001
Religiosity	.57 [.37; .72]	5	5,750	NA	.0024
AUD	.24 [.09; .38]	3	5,162	721	.0221
Drinking quantity	.41 [.11; .64]	6	2,270	NA	.0178
Smoking cessation	.54 [.31; .72]	4	3,613	1,426	.0066
Smoking initiation	.37 [.30; .43]	12	87,253	13,469	< .0001
Smoking quantity	.24 [.14; .34]	6	4,701	NA	.0020
Smoking status	.46 [.35; .56]	15	168,404	20, 584	< .0001
SUD	.29 [.29, .30]	3	1,533,956	241,817	< .0001
Agreeableness	.11 [.05; .18]	11	10,347	NA	.0035
Conscientiousness	.16 [.10; .23]	11	10,347	NA	.0003
Extraversion	.08 [.05; .11]	29	22,483	NA	< .0001
Neuroticism	.10 [.07; .13]	30	23,154	NA	< .0001
Openness	.21 [.14; .28]	11	10,483	NA	< .0001
Body mass index	.16 [.12; .19]	31	131,079	NA	< .0001
Height	.23 [.21; .26]	74	299,763	NA	< .0001
Waist-to-hip ratio	.16 [.08; .24]	5	83,630	NA	.0050
Depression	.14 [.11; .17]	7	1,483,486	211,154	< .0001
Diabetes	.15 [.07; .23]	7	178,522	17,530	.0038

GAD	.14 [.04; .24]	6	116,911	5,284	.018

538

539

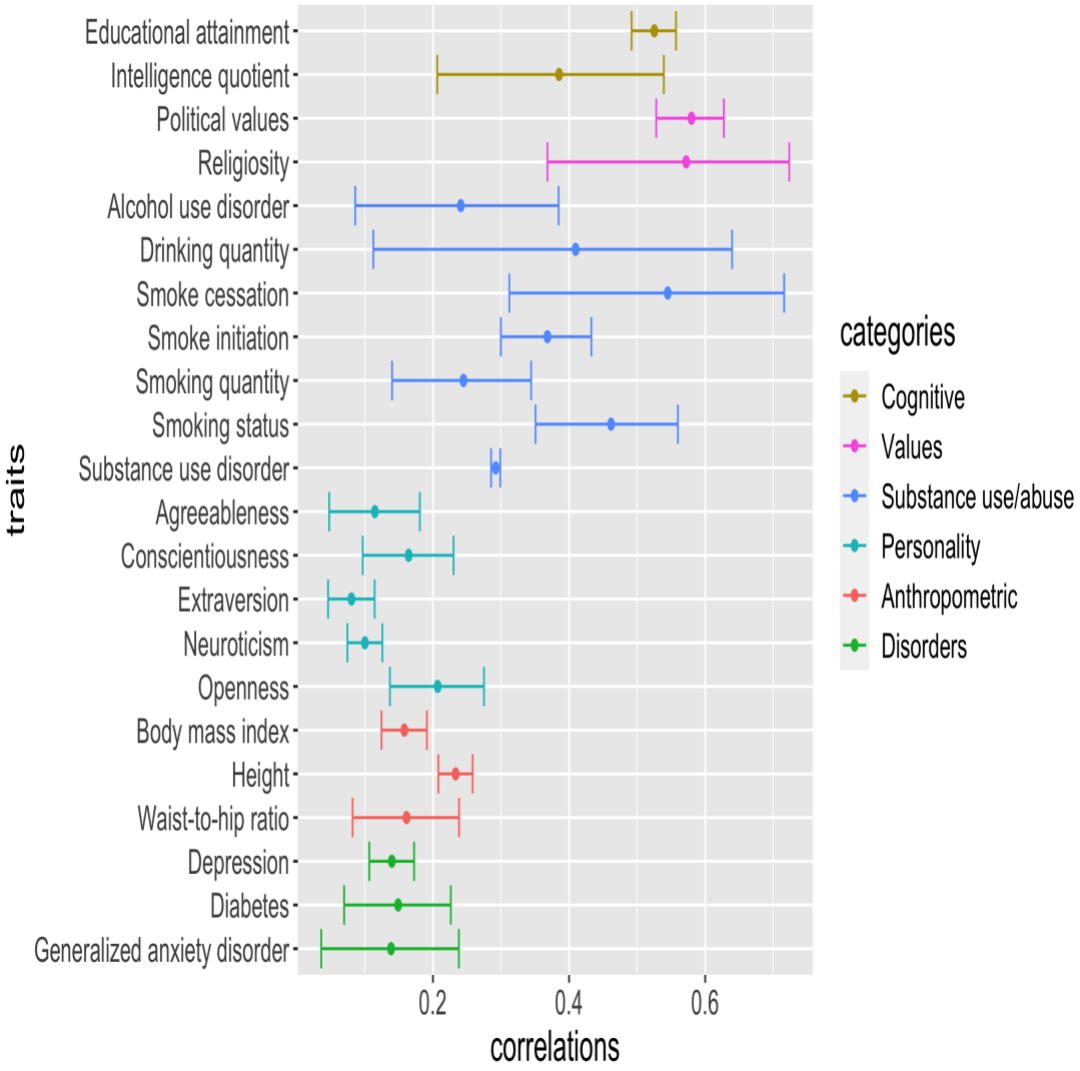
540 **Table 1.** *r* = meta-analyzed random effects spousal correlation (Pearson's *r* for continuous
541 traits; tetrachoric *r* for dichotomous traits), CI = confidence interval, *K* = number of samples
542 meta-analyzed, *N* = number of total spouse pairs meta-analyzed; EA = educational
543 attainment, IQ = intelligence quotient, AUD = alcohol use disorder, SUD = substance use
544 disorder, GAD = generalized anxiety disorder; *Effective N* = $\frac{1-r^2}{se^2}$ + 2 (rearranged from the
545 formula for the standard error estimate).
546
547
548

Trait	I^2 **[CI]**	τ	τ^2 **[CI]**	**Prediction Interval**
EA	93% [91%; 94%]	.100	0.0100 [0.0058; 0.0238]	[0.3568; 0.6607]
IQ	91% [86%; 95%]	.260	0.0675 [0.0288; 0.2524]	[-0.2220; 0.7772]
Political values	80% [62%; 89%]	.082	0.0067 [0.0018; 0.0343]	[0.4256; 0.7014]
Religiosity	95% [91%; 97%]	.204	0.0417 [0.0128; 0.3736]	[-0.0662; 0.8782]
AUD	0% [0%; 90%]	.000	0 [0.0000; 0.3788]	[-0.2221; 0.6153]
Drinking quantity	92% [86%; 96%]	.294	0.0862 [0.0301; 0.5821]	[-0.4228; 0.8671]
Smoking cessation	90% [77%; 96%]	.169	0.0285 [0.0069; 0.4410]	[-0.2102; 0.8928]
Smoking initiation	95% [93%; 97%]	.104	0.0108 [0.0046; 0.0355]	[0.1408; 0.5587]
Smoking quantity	68% [24%; 87%]	.084	0.0070 [0.0006; 0.0642]	[-0.0103; 0.4700]
Smoking status	98% [98%; 99%]	.227	0.0517 [0.0247; 0.1400]	[-0.0095; 0.7651]
SUD	0% [0%; 90%]	.000	0 [0.0000; 0.0404]	[0.2722; 0.3119]

Agreeableness	88% [80%; 93%]	.086	0.0074 [0.0022; 0.0278]	[-0.0908; 0.3108]
Conscientiousness	90% [84%; 94%]	.093	0.0087 [0.0028; 0.0266]	[-0.0564; 0.3698]
Extraversion	68% [54%; 79%]	.068	0.0046 [0.0017; 0.0117]	[-0.0625; 0.2198]
Neuroticism	58% [37%; 72%]	.040	0.0016 [0.0004; 0.0073]	[0.0142; 0.1845]
Openness	87% [78%; 92%]	.090	0.0081 [0.0027; 0.0345]	[-0.0070; 0.4027]
Body mass index	96% [95%; 97%]	.086	0.0074 [0.0038; 0.0129]	[-0.0205; 0.3267]
Height	91% [89%; 92%]	.098	0.0096 [0.0069; 0.0167]	[0.0408; 0.4091]
Waist-to-hip ratio	68% [18%; 88%]	.052	0.0027 [0.0001; 0.0380]	[-0.0265; 0.3380]
Depression	55% [0%; 81%]	.022	0.0005 [0.0000; 0.0085]	[0.0728; 0.2052]
Diabetes	78% [55%; 90%]	.072	0.0052 [0.0005; 0.0445]	[-0.0531; 0.3391]
GAD	51% [0%; 80%]	.076	0.0058 [0.0000; 0.0734]	[-0.0987; 0.3607]

Table 2. Heterogeneity statistics for each trait's meta-analysis. CI = confidence interval, I^2 = Higgins & Thompson's I^2 statistic, a measure of between-study heterogeneity, τ = the estimated standard deviation of the true effect size, τ^2 = the estimated variance of the true effect size; EA = educational attainment, IQ = intelligence quotient, AUD = alcohol use disorder, SUD = substance use disorder, GAD = generalized anxiety disorder.

561 **Figure 1**
562



563
564 The meta-analyzed random effects spousal correlations and 95% confidence intervals for each trait.