

# Assortative Mating on Educational Attainment Leads to Genetic Spousal Resemblance for Causal Alleles

## Abstract:

We examined whether assortative mating for educational attainment (“like marries like”) can be detected in the genomes of ~1,600 UK spouse pairs of European descent. Assortative mating on heritable traits like educational attainment increases the genetic variance and heritability of the trait in the population, which may increase social inequalities. We test for genetic assortative mating in the UK on educational attainment, a phenotype that is indicative of socio-economic status and has shown substantial levels of assortative mating. We use genome-wide allelic effect sizes from a large genome-wide association study on educational attainment (N~300k) to create polygenic scores that are predictive of educational attainment in our independent sample ( $r=.23$ ,  $p<2\times 10^{-16}$ ). The polygenic scores significantly predict partners' educational outcome ( $r=.14$ ,  $p=4\times 10^{-8}$  and  $r=.19$ ,  $p=2\times 10^{-14}$ , for prediction from males to females and vice versa, respectively), and are themselves significantly correlated between spouses ( $r=.11$ ,  $p=7\times 10^{-6}$ ). Our findings provide molecular genetic evidence for genetic assortative mating on education in the UK.

**Keywords:** Assortative Mating; Educational Attainment; Polygenic Scores

## 1. Introduction

Humans generally do not choose their mates randomly. In search for a suitable mate, among the highest-ranking qualities people look for in a potential partner are intelligence and educational attainment (Buss & Barnes, 1986; Zietsch, Verweij, & Burri, 2012). Previous work consistently shows substantial assortative mating for intelligence and educational attainment, with spousal correlations for intelligence ranging between .33 and .72 (Bouchard & McGue, 1981; Gualtieri, 2013; Mascie-Taylor & Vandenberg, 1988; Watson, et al., 2004) and for educational attainment between .45 and .66 (Abdellaoui, et al., 2015; Conley, et al., 2016; Watson, et al., 2004; Zietsch, Verweij, Heath, & Martin, 2011). Assortative mating can occur via different mechanisms (which are not always mutually exclusive). Partners can become more similar to each other over the course of their relationship (i.e., convergence); however, there is no evidence for convergence for cognitive abilities and educational attainment (Mascie-Taylor & Vandenberg, 1988; Watson, et al., 2004; Zietsch, et al., 2011). This suggests that assortative mating for educational attainment is due to initial partner choice. This can happen because of social homogamy, where similar people find themselves in similar social environments because of their social background, and/or because of phenotypic matching, where people select their partner based on similarity in characteristics.

The consequences of assortative mating on education and cognitive abilities are relevant for society and for the genetic make-up and therefore the evolutionary development of subsequent generations (Thiessen & Gregg, 1980). Assortative mating increases the variance of characteristics in the population, and may increase social inequality with respect to education or income (Schwartz, 2013). Greenwood et al. (2014) for instance reported a rise in assortative mating for educational attainment in the United States between 1960 and 2005 and showed that this clustering of academic success may have caused an increase in income inequality. It is *a priori* very plausible that phenotypic similarity between partners on heritable traits is reflected in their genomic similarities, and thus in the genetic composition of their offspring. Assortative mating on a heritable trait increases the additive genetic variance for genetic loci associated with that trait, as well as for other traits that are genetically correlated with it (Crow & Felsenstein, 1968; Fisher, 1918; Lande, 1977), as assortative mating generates phenotypes with more extreme genetic values. Consistent with the increase in assortment for educational attainment (Greenwood, et al., 2014; Schwartz, 2013), this may explain why heritability estimates for educational attainment have risen over time (Branigan, McCallum, & Freese, 2013), although there may also be other explanations for this increase, such as the recently increased equality in educational opportunities (Colodro-Conde, Rijdsdijk, Tornero-Gómez, Sánchez-Romera, & Ordoñana, 2015). Another genetic consequence of assortative mating on education is the influence on genome-wide ancestral variation and homozygosity. Abdellaoui et al. (2015) showed that more educated individuals are more likely to migrate, which increases their chance of meeting a spouse with a different ancestral background. Accordingly, assortment on educational attainment can result in greater ancestral variation and lower levels of genome-wide

homozygosity, a genetic signature used to study effects of inbreeding, within the offspring of higher educated spouse pairs.

Several studies have tried to detect assortative mating on a molecular genetic level by estimating spousal resemblance on genome-wide single nucleotide polymorphisms (SNPs) (Domingue, Fletcher, Conley, & Boardman, 2014; Guo, Wang, Liu, & Randall, 2014; Sebro, Hoffman, Lange, Rogus, & Risch, 2010). These studies report spouses to be more similar on genome-wide SNPs than expected under random mating. However, these reported spousal resemblances are more likely to be explained by population stratification, i.e., spouse pairs sharing more ancestry than random male-female pairs (Abdellaoui, Verweij, & Zietsch, 2014; Sebro, et al., 2010), than by phenotypic assortative mating. Assortative mating on complex phenotypes, such as education, intelligence, personality, psychiatry, or height, is expected to lead to genetic spousal resemblance. However, these traits are influenced by many genetic variants throughout the genome with very small individual effects that require exceptionally large sample sizes to detect. The largest patterns of genome-wide variation, which can be captured with principal component analyses (PCA) in much smaller datasets, reflect ancestry differences (Price, et al., 2006), correlate strongly with geography and show significant spouse correlations (Abdellaoui, Hottenga, de Knijff, et al., 2013). Geographic proximity is a strong predictor of shared ancestry and a major determinant of potential spouse pairs meeting, especially in the presence of social catalysts that narrow mate choice and correlate with geography, such as religion (Abdellaoui, Hottenga, Xiao, et al., 2013; Haber, et al., 2013). We therefore expect spousal resemblance on a genome-wide level to be dominated by shared ancestry, and indeed the above studies do not show a significant genetic spousal resemblance once ancestry is appropriately accounted for. A trait-based approach is more powerful, less susceptible to population stratification, and thus more informative in detecting genetic assortative mating than estimating allelic spousal resemblance in a hypothesis-free manner. With the advent of large-scale genome-wide association studies (GWASs), we can now quantify significant portions of a person's genetic predisposition for a wide range of traits with polygenic scores by summing their individual alleles weighted by their estimated effect sizes. Polygenic scores can have significant predictive power and generally improve for complex traits when adding SNPs that individually did not reach genome-wide significance (Dudbridge, 2013).

The highly polygenic trait educational attainment is well suited for a study on genetic assortative mating because the phenotype itself is subject to high levels of assortment and genome-wide estimates of allelic effect sizes are available from large GWASs. Conley et al. (2016) show that polygenic scores based on results from a GWAS on educational attainment of ~126,000 subjects (Rietveld, et al., 2013) significantly correlate between spouse pairs born between 1920 and 1950 in the US. We use genome-wide effect sizes from a GWAS on educational attainment of ~300,000 subjects (Okbay, et al., 2016) to create polygenic scores for couples born between 1919 and 1994 from the UK Household Longitudinal Study (UKHLS), a survey that aims to be representative of the UK population. Given similar levels of phenotypic

assortative mating in the US and the UK, we expect to replicate that there is genetic assortative mating for educational attainment and find higher levels of genetic assortative mating than Conley et al. (2016) given the more powerful summary statistics and a novel and more powerful polygenic score approach (Vilhjálmsón, et al., 2015). We test whether individuals' polygenic risk scores for educational attainment can predict their partners' educational attainment, and their partners' polygenic scores. We control for similarities in ancestral background by taking into account ancestry-informative principal components (PCs).

## 2. Materials and Methods

### 2.1. Phenotypes

The sample is derived from the UK Household Longitudinal Study: *Understanding Society* (UKHLS) (Buck & McFall, 2011), a representative sample of the UK population. 9,944 individuals were genotyped, including 1,699 pairs who were living together either as husband and wife or as a couple. Individuals under 25 years of age were removed from the analyses, because they are likely to not have reached their final education level; this resulted in an N of 8,989. For the cross-spouse analyses we also removed all pairs where either partner was under 25, resulting in a sample of 1,616 spouse pairs.

We derived a variable for individuals' educational attainment as follows: 0=no educational qualifications; 1=GCSE (national exams taken at age 16) or "other qualifications"; 2=A-level or equivalent (national exams taken at age 18, roughly equivalent to French Baccalaureate or US High School Diploma); 3=University degree or equivalent. Educational attainment was standardized to have a mean of 0 and a standard deviation of 1.

The UKHLS is a stratified probability sample of the UK population. The dataset for the nurse visit sample (from which the SNP data are derived) includes response weights which are meant to account for ascertainment bias and non-response, including non-participation in the nurse visit and not donating blood. We used the cross-sectional weights, i.e., the reciprocal of the probability of blood measures to be present for a particular individual, predicted from a variety of socio-economic characteristics. Further details are given in reference (Benzeval, Davillas, Kumari, & Lynn, 2014). For analyses where each case represents a pair of partners, such as the main regressions on partner characteristics, we used the arithmetic mean of male and female partner's weight.

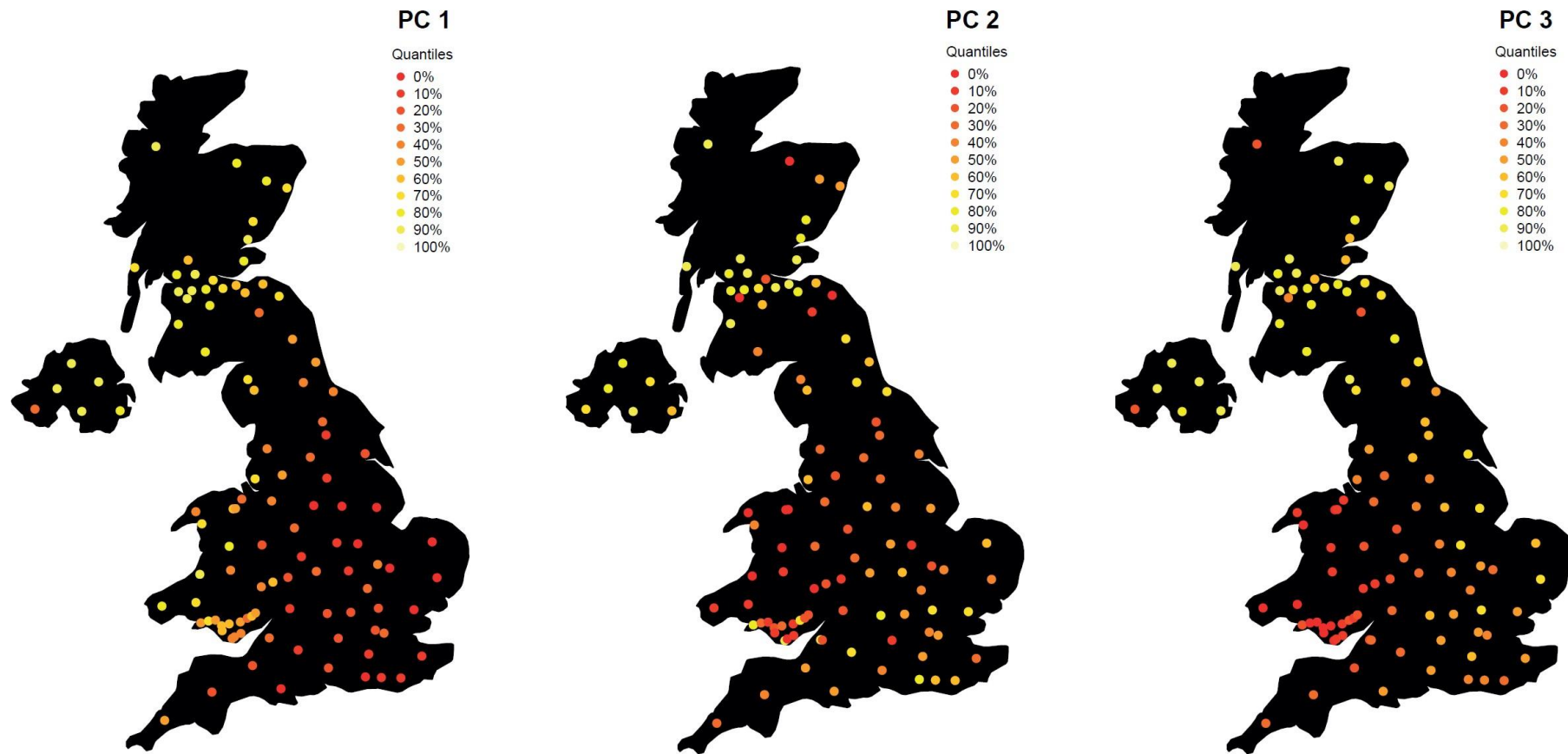
### 2.2. Genotyping, Quality Control (QC), and Principal Component Analysis (PCA)

Genotyping was done on the Illumina HumanCoreExome chip for White/European participants of Waves 2 and 3 of the *Understanding Society* study. QC was performed on the entire set of 9,944 subjects in PLINK (Purcell, et al., 2007), and only autosomal SNPs were included. SNPs were excluded if they: 1) had a missing rate greater than 5%; 2) showed a minor allele frequency (MAF) smaller than 5%; 3) deviated from Hardy–Weinberg equilibrium (HWE) with a  $p$ -value smaller than  $10^{-8}$ . The QC resulted in 261,965 SNPs with a mean individual genotyping rate of >99.9% (ranging from 97.2% to 99.99%, with only 15 individuals having >1% missingness). There were no individuals detected with a non-European or non-British ancestry by projecting principal components (PCs) from the 1000 Genomes dataset (procedure described in more detail in the supplementary material of reference (Abdellaoui, Hottenga, de Knijff, et al., 2013)). To control for ancestry differences within the UK, we conducted a PCA on

the genotype data in EIGENSTRAT (Price, et al., 2006). In order to detect the relatively small ancestry differences within the UK, we pruned for linkage disequilibrium (LD) (window size=50, number of SNPs to shift after each step=5, based on a variance inflation factor [VIF] of 2) and removed long-range LD regions, since LD can result in larger patterns of variation than ancestry differences within relatively homogeneous populations (Abdellaoui, Hottenga, de Knijff, et al., 2013). After minimizing LD, 91,708 autosomal SNPs remained. The PCA was conducted on unrelated individuals (9,091 out of 9,944 subjects) and projected onto the rest. Unrelated individuals were chosen using GCTA (Yang, Lee, Goddard, & Visscher, 2011), by excluding one of each pair of individuals with an estimated genetic relationship of  $> .025$  (i.e., closer related than third or fourth cousin).

### 2.3. Polygenic Scores

Polygenic scores were computed using LDpred (Vilhjalmsson, et al., 2015), a recently developed method that creates an unbiased predictor with increased accuracy by conditioning on a genetic architecture prior and linkage disequilibrium (LD) information from a reference sample. Risk scores in our sample were generated with effect sizes obtained from the latest educational attainment GWAS (Okbay, et al., 2016) and the LD information from the European populations in the 1000 Genomes reference set (in our case: Utah Residents (CEPH) with Northern and Western European Ancestry, Finnish, British, Iberian, and Toscani individuals,  $N=381$ ). Vilhjalmsson et al. (2015) showed theoretically, with simulations, and empirically that this method outperforms the traditional pruning/thresholding approach. We varied the expected fraction of causal markers (.01%, .03%, .1%, .3%, 1%, 3%, 10%, 30%, 50%, 75%, 100%) in order to optimize the prediction accuracy (Vilhjalmsson, et al., 2015), similar to varying  $p$ -values in order to determine which SNPs are included in traditional polygenic score approaches. Polygenic scores were standardized to have mean of 0 and a standard deviation of 1.



*Figure 1: The first three principal components (PCs) that show significant correlations with geography and significant spouse correlations. The mean PC value per county was computed, divided into 10 percentile groups, and plotted.*

### 3. Results

A PCA was conducted on the 261,965 SNPs that remain after QC. The top 3 PCs show significant and substantial spouse correlations as well as significant correlations with geography (Table 1 and Figure 1), implying that these three PCs capture ancestry differences within the UK population. The polygenic scores for education analyzed below are residualized by removing the effects of the first 10 PCs.

*Table 1: Spouse correlations for the top ten genetic principal components (PCs), and correlations between PCs and latitude, longitude, and the educational attainment (EA) polygenic score (p-values between brackets).*

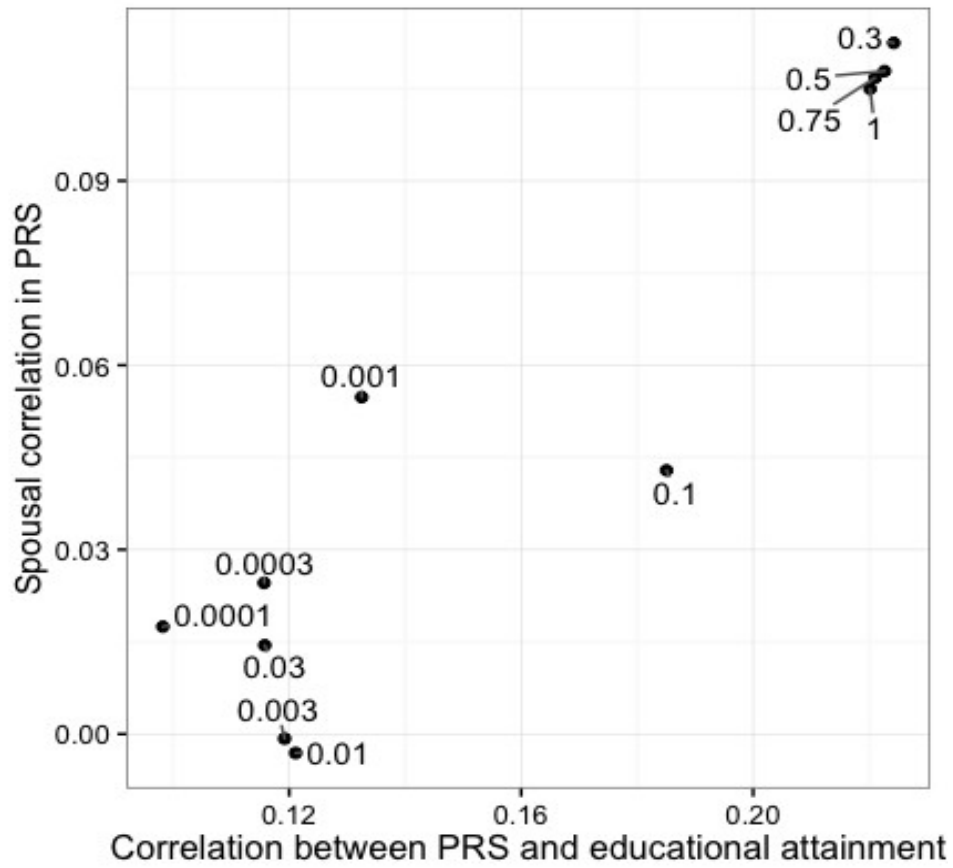
PC	Spouse Correlation	Latitude	Longitude	EA Polygenic Score
1	.33 ( $<2 \times 10^{-16}$ )	-.49 ( $<2 \times 10^{-16}$ )	.41 ( $<2 \times 10^{-16}$ )	.04 (.02)
2	.60 ( $<2 \times 10^{-16}$ )	-.03 (.31)	.14 ( $4 \times 10^{-6}$ )	-.01 (.36)
3	.30 ( $<2 \times 10^{-16}$ )	.11 ( $1.4 \times 10^{-6}$ )	.23 ( $<2 \times 10^{-16}$ )	.005 (.80)
4	.005 (.87)	-.12 ( $6.6 \times 10^{-8}$ )	.10 ( $2 \times 10^{-6}$ )	.02 (.24)
5	.01 (.66)	-.01 (.66)	.002 (.92)	-.02 (.31)
6	.01 (.58)	-.06 (.01)	.09 ( $2 \times 10^{-5}$ )	.001 (.94)
7	.01 (.57)	-.02 (.26)	.06 (.004)	.003 (.87)
8	.05 (.05)	-.04 (.10)	.03 (.15)	.02 (.15)
9	-.01 (.57)	-.05 (.03)	.05 (.03)	.02 (.22)
10	.01 (.58)	-.02 (.40)	-.02 (.40)	.02 (.24)

Consistent with previous studies, couples in our sample show significant assortative mating for educational attainment, with a Spearman's rank correlation between spouses of .45 ( $p < 2 \times 10^{-16}$ ;  $N=1,604$  pairs). Polygenic scores that assume the fraction of causal markers to be 30% or higher show the strongest correlations with educational attainment ( $r=.22$ ,  $p < 2 \times 10^{-16}$ ,  $N=8,982$ ), consistent with education being a highly polygenic trait (Figure 2 and Figure 3). As Figure 3 shows, this polygenic score is highly predictive of educational attainment; subjects with polygenic scores in the highest quintile for instance were more than twice as likely to go to university as those with scores in the lowest quintile. We use the fraction of 30% for subsequent analyses. Polygenic scores for males significantly predict their partners' educational outcome ( $r=.14$ ,  $p=4 \times 10^{-8}$ ,  $N=1,608$ ) and female polygenic scores significantly predict their partner's education ( $r=.19$ ,  $p=2 \times 10^{-14}$ ,  $N=1,609$ ). Results are barely affected by controlling for partner's age (males:  $r=.16$ ,  $p=2 \times 10^{-12}$ ; females:  $r=.18$ ,  $p=5 \times 10^{-14}$ ). The polygenic scores themselves also show significant spouse correlations ( $r=.11$ ,  $p=7 \times 10^{-6}$ ,  $N=1,616$ ). Without residualizing

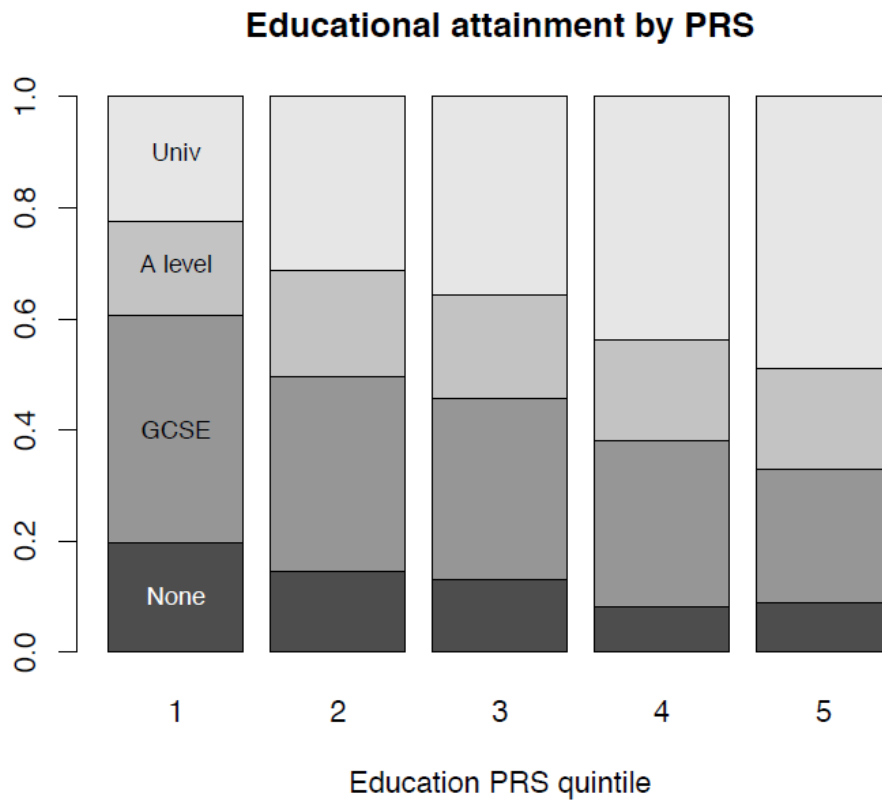


polygenic scores for the first 10 PCs, spousal correlations were practically unchanged ( $r=.11$ ,  $p=7\times 10^{-6}$ ,  $N=1,616$ ). When dropping the cross-sectional weights (see Materials and Methods), spouse correlations were slightly smaller but remained significant (male polygenic score with female education:  $r=.14$ ,  $p=3\times 10^{-8}$ ,  $N=1,668$ ; female polygenic score with male education:  $r=.18$ ,  $p=5\times 10^{-14}$ ,  $N=1,668$ ; male polygenic score with female polygenic score:  $r=.09$ ,  $p=3\times 10^{-4}$ ,  $N=1,678$ ). The spouse correlations of the polygenic scores were small and non-significant when assuming a lower fraction of causal markers (<30%), as would be expected when they are less accurate estimates of someone's genetic predisposition for education (Figure 2). When only including spouse pairs with offspring ( $N=621$ ), the spouse correlations become slightly stronger and remain significant (male polygenic score with female education:  $r=.17$ ,  $p=4\times 10^{-6}$ ,  $N=620$ ; female polygenic score with male education:  $r=.22$ ,  $p=1\times 10^{-8}$ ,  $N=619$ ; male polygenic score with female polygenic score:  $r=.15$ ,  $p=1\times 10^{-4}$ ,  $N=621$ ). Results are also robust to using alternative coding of educational attainment: 1) using a 5-level scoring whereby postgraduate degrees are coded as educational level 4: male polygenic score with female education  $r=.17$ ,  $p=2\times 10^{-8}$ ,  $N=1605$ , female polygenic score with male education:  $r=.22$ ,  $p=7.3\times 10^{-14}$ ,  $N=1594$ ; 2) dropping subjects with vocational qualifications: male polygenic score with female education  $r=.14$ ,  $p=3.5\times 10^{-8}$ ,  $N=1368$ , female polygenic score with male education:  $r=.20$ ,  $p=2.1\times 10^{-14}$ ,  $N=1369$ .

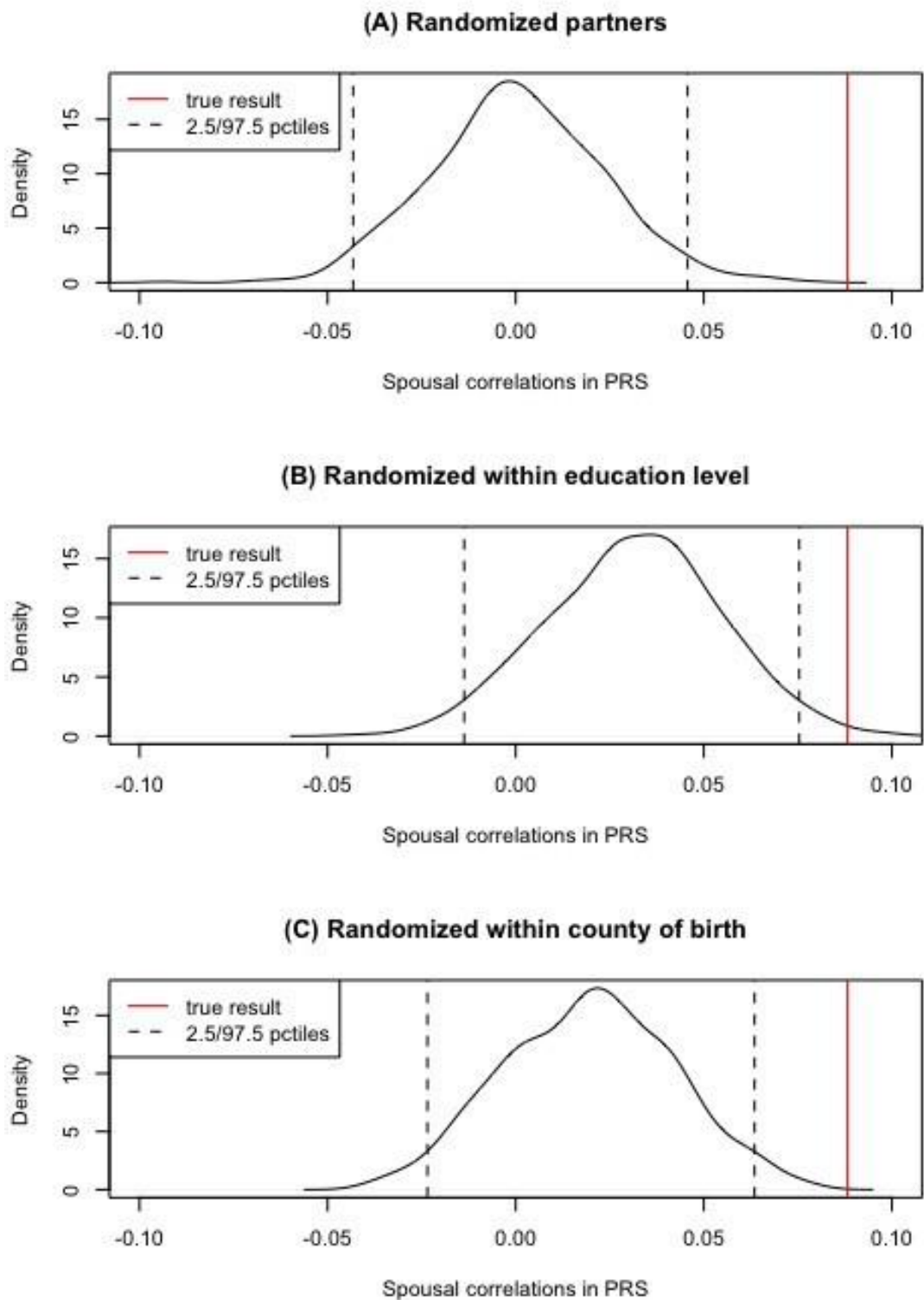
To test whether our data can be explained by random mating within specific subpopulations, we performed permutation tests within the whole sample, within educational levels, and within birth counties. We re-matched individuals with random "partners" 1000 times, creating a distribution of random test statistics (i.e., the coefficient of the spouse correlation of the polygenic scores) under the null hypothesis. We then compare the observed test statistic with respect to this distribution. Figure 4A shows the results within a permuted sample. Figure 4B shows the results within a permuted sample conditional on educational levels, i.e., partners were matched randomly within the same educational level. Figure 4C shows results within a permuted sample conditional on geographical location, i.e., partners were matched according to county of the UK. In all three cases the observed set of spouse pairs showed greater similarities in polygenic scores than 97.5% of the randomly generated sets of spouse pairs. Thus, there was significant genetic assortative mating for educational attainment irrespective of educational level and geographic location.



**Figure 2:** The correlations between the polygenic score (PRS) and educational attainment on the x-axis and the spouse correlations for PRS on the y-axis both increase as the fraction of causal markers for the polygenic score increases beyond .3. Labels show assumed fraction of causal markers..



**Figure 3:** The proportion of each of the four educational levels (Univ = University degree or equivalent; A level = A-level or equivalent (national exams taken at age 18); GCSE = General Certificate of Secondary Education (national exams taken at age 16); None = no educational qualifications) per quintile of the polygenic scores (PRS) for educational attainment.



**Figure 4:** Three permutation procedures. Plots show the density of spousal PRS correlations under the null hypotheses that spouses are chosen randomly (A) within the whole sample, (B) among people of the same education level, and (C) among people of the same birth county.

#### 4. Discussion

This study provides empirical evidence for genetic assortative mating in the UK on a cognitive and behavioral trait. We show that this assortment has consequences on a genetic and thus a biological level. The polygenic scores significantly predicted partners' educational outcome (for both sexes), i.e., individuals with a stronger genetic predisposition for higher educational attainment have partners who are more educated. Also, the educational attainment polygenic scores themselves are correlated between spouses, which is strong evidence for the presence of genetic assortative mating for education in the UK. Within counties and within educational levels, spouse pairs still resembled each other with respect to their polygenic score. Polygenic scores for educational attainment explain 4.8% of the educational outcome in our sample, similar to the explained variance reported in the original GWAS (Okbay, et al., 2016). Since the polygenic scores only partly reflect the genetic predisposition for educational attainment, and yet show a relatively strong spouse correlation compared to the phenotypic spouse correlation, we can assume that a considerable part of assortative mating on educational attainment is genetic. The spouse correlations of the polygenic scores ( $\sim .11$ ) are in a similar range to what Conley et al. (2016) reported ( $\sim .13$ ), as were their phenotypic spouse correlations (.53). Conley et al. (2016) used summary statistics from a smaller GWAS (Rietveld, et al., 2013) that capture less of the individual differences for educational attainment ( $\sim 2\%$  in the original GWAS from Rietveld et al. (2013), 1.7% in our UK sample, and 3.2% in the Conley et al. (2016) study) and a more traditional polygenic score approach (i.e., an approach that did not take LD structure into account). When repeating the analysis in our UK sample using summary statistics from Rietveld et al. (2013), the spouse correlation decreases to .06 ( $p = .01$ ). It is not certain whether the differences between the UK sample and the US sample are due sample fluctuation, other (statistical) artefacts, or actual differences in the degree of (genetic) assortative mating between the UK and the US, but our results do provide a solid replication of the presence of genetic assortative mating in contemporary Western societies.

Assortative mating can result in biased heritability estimates in twin studies when not accounted for. Twin studies disentangle variance in traits into genetic, shared environmental, and residual factors by comparing similarities within identical and non-identical twins. The assumption of twin models is that identical twins share all their genes and non-identical twins share 50% of their segregated genes. Assortative mating increases the genetic similarity of non-identical twins above the assumed 50%. If assortative mating is not explicitly modelled in the twin model (which generally is not the case), the increased resemblance between non-identical twins due to assortative mating will result in an over-estimation of the shared environmental component of variance, and under-estimation of the heritable component. The substantial levels of assortative mating for educational attainment imply that the genetic influences may have been underestimated in many twin studies. In a meta-analysis of virtually all twin studies on educational attainment in the last 50 years ( $N=31$ ), Polderman et al. (2015) reported a genetic

and shared environmental estimate for educational attainment of .52 and .27, respectively. However, assuming a spousal correlation of approximately .5, the shared environmental component can be entirely attributed to assortative mating. When correcting the variance component estimates for educational attainment with a procedure described by Martin (1978), the genetic component increases to ~80% while the shared environmental component drops to zero<sup>1</sup>. Interestingly, Polderman et al. observe the lowest heritability estimates for traits in the categories ‘environment’ (including education), reproduction, and social values domains. These include traits for which particularly high levels of assortative mating have been observed, suggesting that the lower heritability estimates for these domains may in part be explained by assortative mating that is not accounted for.

Social inequality has been around in many historical and modern human societies, but is considered, as President Obama recently put it (Obama, 2013), as “the defining challenge of our time”. Assortative mating on heritable traits that are indicative of socio-economic status, such as educational attainment, increases genetic variance in such a way that the inequality in genetic capital grows. When growing social inequality is (partly) driven by a growing biological inequality, inequalities in society may be harder to overcome. Effects of assortative mating may accumulate with each generation. The increasing social mobility for females during the second half of the 20<sup>th</sup> century possibly also led to an increase in assortative mating as well as an increase in social inequality (Greenwood, et al., 2014; Schwartz, 2013). Conley et al. (2016) recently showed that phenotypic assortative mating for educational attainment has increased in the US for ~2,000 spouse pairs born between 1920 and 1955, but did not observe an increase in genetic assortative mating based on polygenic scores from a GWAS on ~126,000 individuals, likely due to a lack of power.

The realization that assortative mating on talents may have genetic consequences is far beyond than a century old (Galton, 1869). Molecular genetics has progressed sufficiently to empirically reveal the presence of genetic assortative mating on a trait that reflects a collection of cognitive and behavioral talents. The presence of genetic assortative mating on a broad socio-economic related trait like educational attainment may have consequences for genetic studies as well as for society.

---

<sup>1</sup> Martin (1978) used the following formula to correct shared environmental influences (C) for assortative mating:  $C_{adj} = \frac{c_R}{h_R^2} \times \frac{1}{1-A}$ , where  $h_R^2$  and  $c_R$  are the genetic and shared environmental influences as estimated by the twin model, and A is the correlation between additive genetic values of mates, which is a function of the observed value for assortative mating ( $\mu$ ) and  $h_R$ :  $A = 0.5 * [1 - \sqrt{1-4\mu h_R}]$ .

## References

- Abdellaoui, A., Hottenga, J.-J., de Knijff, P., Nivard, M. G., Xiao, X., Scheet, P., Brooks, A., Ehli, E. A., Hu, Y., & Davies, G. E. (2013). Population structure, migration, and diversifying selection in the Netherlands. *European journal of human genetics*, *21*, 1277-1285.
- Abdellaoui, A., Hottenga, J.-J., Xiao, X., Scheet, P., Ehli, E. A., Davies, G. E., Hudziak, J. J., Smit, D. J., Bartels, M., & Willemsen, G. (2013). Association between autozygosity and major depression: stratification due to religious assortment. *Behav Genet*, *43*, 455-467.
- Abdellaoui, A., Hottenga, J. J., Willemsen, G., Bartels, M., van Beijsterveldt, T., Ehli, E. A., Davies, G. E., Brooks, A., Sullivan, P. F., Penninx, B. W., de Geus, E. J., & Boomsma, D. I. (2015). Educational attainment influences levels of homozygosity through migration and assortative mating. *PLoS One*, *10*, e0118935.
- Abdellaoui, A., Verweij, K. J., & Zietsch, B. P. (2014). No evidence for genetic assortative mating beyond that due to population stratification. *Proc Natl Acad Sci U S A*, *111*, E4137.
- Benzeval, M., Davillas, A., Kumari, M., & Lynn, P. (2014). Understanding Society: UK Household Longitudinal Study: Biomarker User Guide and Glossary. *Colchester: University of Essex*.
- Bouchard, T. J., & McGue, M. (1981). Familial studies of intelligence: A review. *Science*, *212*, 1055-1059.
- Branigan, A. R., McCallum, K. J., & Freese, J. (2013). Variation in the heritability of educational attainment: An international meta-analysis. *Social forces*, *92*, 109-140.
- Buck, N., & McFall, S. (2011). Understanding Society: design overview. *Longitudinal and Life Course Studies*, *3*, 5-17.
- Buss, D. M., & Barnes, M. (1986). Preferences in human mate selection. *Journal of Personality and Social Psychology*, *50*, 559-570.
- Colodro-Conde, L., Rijdsdijk, F., Tornero-Gómez, M. J., Sánchez-Romera, J. F., & Ordoñana, J. R. (2015). Equality in Educational Policy and the Heritability of Educational Attainment. *PLoS One*, *10*, e0143796.
- Conley, D., Laidley, T., Belsky, D. W., Fletcher, J. M., Boardman, J. D., & Domingue, B. W. (2016). Assortative mating and differential fertility by phenotype and genotype across the 20th century. *Proceedings of the National Academy of Sciences*, 201523592.
- Crow, J. F., & Felsenstein, J. (1968). The effect of assortative mating on the genetic composition of a population. *Eugenics Quarterly*, *15*, 85-97.
- Domingue, B. W., Fletcher, J., Conley, D., & Boardman, J. D. (2014). Genetic and educational assortative mating among US adults. *Proceedings of the National Academy of Sciences*, *111*, 7996-8000.
- Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *PLoS Genet*, *9*, e1003348.
- Fisher, R. (1918). The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Philosophical Transactions of the Royal Society of Edinburgh*.
- Galton, F. (1869). *Hereditary genius*: Macmillan and Company.
- Greenwood, J., Guner, N., Kocharkov, G., & Santos, C. (2014). Marry your like: Assortative mating and income inequality. In: National Bureau of Economic Research.
- Gualtieri, C. T. (2013). Husband-wife correlations in neurocognitive test performance. *Psychology*, *4*, 771.
- Guo, G., Wang, L., Liu, H., & Randall, T. (2014). Genomic assortative mating in marriages in the United States. *PLoS One*, *9*, e112322.

- Haber, M., Gauguier, D., Youhanna, S., Patterson, N., Moorjani, P., Botigué, L. R., Platt, D. E., Matisoo-Smith, E., Soria-Hernanz, D. F., & Wells, R. S. (2013). Genome-wide diversity in the levant reveals recent structuring by culture. *PLoS Genet*, *9*, e1003316.
- Lande, R. (1977). The influence of the mating system on the maintenance of genetic variability in polygenic characters. *Genetics*, *86*, 485-498.
- Martin, N. G. (1978). Genetics of sexual and social attitudes in twins. *Twin Research: Psychology and Methodology*, *24*, 13-23.
- Mascie-Taylor, C. G., & Vandenberg, S. G. (1988). Assortative mating for IQ and personality due to propinquity and personal preference. *Behav Genet*, *18*, 339-345.
- Obama, B. (2013). Remarks by the president on economic mobility. *The White House, Office of the Press Secretary*, <http://www.whitehouse.gov/the-press-office/2013/12/04/remarks-president-economic-mobility>.
- Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., Turley, P., Chen, G.-B., Emilsson, V., & Meddens, S. F. W. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*.
- Polderman, T. J., Benyamin, B., De Leeuw, C. A., Sullivan, P. F., Van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet*.
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*, *38*, 904-909.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., De Bakker, P. I., & Daly, M. J. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*, 559-575.
- Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., Westra, H.-J., Shakhbazov, K., Abdellaoui, A., & Agrawal, A. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science*, *340*, 1467-1471.
- Schwartz, C. R. (2013). Trends and variation in assortative mating: Causes and consequences. *Annual Review of Sociology*, *39*, 451-470.
- Sebro, R., Hoffman, T. J., Lange, C., Rogus, J. J., & Risch, N. J. (2010). Testing for non-random mating: evidence for ancestry-related assortative mating in the Framingham heart study. *Genetic epidemiology*, *34*, 674-679.
- Thiessen, D., & Gregg, B. (1980). Human assortative mating and genetic equilibrium: An evolutionary perspective. *Ethology and Sociobiology*, *1*, 111-140.
- Vilhjálmsón, B. J., Yang, J., Finucane, H. K., Gusev, A., Lindström, S., Ripke, S., Genovese, G., Loh, P.-R., Bhatia, G., & Do, R. (2015). Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *The American Journal of Human Genetics*, *97*, 576-592.
- Watson, D., Klohnen, E. C., Casillas, A., Simms, E. N., Haig, J., & Berry, D. S. (2004). Match makers and deal breakers: Analyses of assortative mating in newlywed couples. *Journal of Personality*, *72*, 1029-1068.
- Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, *88*, 76-82.
- Zietsch, B. P., Verweij, K. J., Heath, A. C., & Martin, N. G. (2011). Variation in human mate choice: simultaneously investigating heritability, parental influence, sexual imprinting, and assortative mating. *Am Nat*, *177*, 605-616.
- Zietsch, B. P., Verweij, K. J. H., & Burri, A. V. (2012). Heritability of preferences for multiple cues of mate quality in humans. *Evolution*, *66*, 1762-1772.



**Acknowledgements:**

Data are available from the UK Household Longitudinal Survey (<http://www.understandingsociety.ac.uk>). Computer code is available from the authors on request.

# **Assortative Mating on Educational Attainment Leads to Genetic Spousal Resemblance for Causal Alleles**

*David Hugh-Jones<sup>1\*</sup>†, Karin J.H. Verweij<sup>2†</sup>, Beate St. Pourcain<sup>3</sup>, Abdel Abdellaoui<sup>2\*</sup>*

<sup>1</sup> *Department of Economics, University of East Anglia, Research Park, Courtyard B, Norwich NR4 7TJ, Norwich, England.*

<sup>2</sup> *Department of Biological Psychology, VU University, van der Boechorstraat 1, 1081 BT Amsterdam, The Netherlands.*

<sup>3</sup> *Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands.*

†Joint first authors

## **\* Correspondence to:**

David Hugh-Jones, e-mail: [davidhughjones@gmail.com](mailto:davidhughjones@gmail.com)

Abdel Abdellaoui, e-mail: [a.abdellaoui@vu.nl](mailto:a.abdellaoui@vu.nl)