



MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Der Einfluss von genetischer Prädisposition, Einkommen, Bildung und Geschlecht auf die Anzahl an Nachkommen in den Vereinigten Staaten“

verfasst von / submitted by

Rafael Steinlesberger, BEd

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Master of Education (MEd)

Wien, 2021 / Vienna 2021

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Betreut von / Supervisor:

UA 199 502 525 02

Masterstudium Lehramt Sek (AB) UF Biologie und
Umweltkunde UF Psychologie und Philosophie

Assoz. Prof. Mag. Dr. Martin Fieder, Privatdoz.

Inhaltsverzeichnis

1. Abstract	3
2. Zusammenfassung.....	3
3. Einleitung.....	3
4. Methoden.....	8
4.1. Datensample	8
4.2. Konstruktion der Polygenic Scores und Variablen	9
4.2.1.PGS Educational attainment (EA)	9
4.2.2.PGS Number of children ever born (NEB)	10
4.2.3.Ausbildungsjahre	10
4.2.4.Einkommensvariablen	11
5. Ergebnisse	12
6. Diskussion.....	14
7. Literaturverzeichnis.....	20
8. Anhang	24

Abstract

Recent findings on a negative phenotypic and genetic association between educational attainment and fertility indicate signs of ongoing natural selection. This study extends on previous work investigating the relative effect size of polygenic scores (PGS) for educational attainment (EA) and number of children ever born (NEB) as well as earnings and actual years of schooling on reproductive success in a sample of 11.850 Americans of European ancestry who participated in the Health and Retirement Study. The results imply that the NEB PGS has the biggest influence on reproductive success, suggesting a strong genetic basis. A rise in years of education is associated with fewer children born for both sexes but slightly fewer for women. As for income, the results show a negative effect on reproductive success for women and on the contrary a positive effect for men. The negative association of the EA PGS found in previous literature could not be replicated under the circumstances of the multivariate regression conducted, showing no significant effect on reproductive success. The strong effect of gene-environment interaction and socioeconomic factors on these findings are discussed in this thesis as well as inconsistencies concerning EA PGS. The found results lead to the conclusion that there is not enough evidence to suggest longterm selection against genetic variants associated with EA.

Zusammenfassung

Neueste Erkenntnisse über einen negativen phänotypischen und genetischen Zusammenhang zwischen Bildungsniveau und Kinderzahl deuten auf eine fortlaufende natürliche Selektion hin. Anknüpfend an bisherige Arbeiten, untersucht diese Studie die relative Effektstärke der polygenic scores (PGS) von Bildungsniveau (EA) und Kinderzahl (NEB), sowie des Einkommens und der tatsächlichen Schuljahre auf den Fortpflanzungserfolg in einer Stichprobe von 11.850 Amerikanern europäischer Abstammung, welche an der Health and Retirement Study teilgenommen haben. Die Ergebnisse zeigen, dass der NEB PGS den größten Einfluss auf den Fortpflanzungserfolg hat, was auf eine starke genetische Grundlage hindeutet. Ein Anstieg des Bildungsniveaus ist bei beiden Geschlechtern mit einer geringeren Kinderzahl verbunden, wobei der Effekt bei Frauen etwas stärker ausfällt. Was das Einkommen anbelangt, so konnte bei Frauen ein negativer Effekt auf den Reproduktionserfolg festgestellt werden, bei Männern hingegen ein positiver Effekt. Die negative Assoziation des EA PGS, welche in früherer Literatur beschrieben wurde, konnte unter den Bedingungen der hier durchgeführten multivariaten Regression nicht repliziert werden, da kein signifikanter Effekt auf den Reproduktionserfolg festgestellt werden konnte. Im Weiteren werden der starke Einfluss von Gen-Umwelt-Interaktionen und sozioökonomischen Faktoren auf diese Ergebnisse, sowie Unstimmigkeiten in Bezug auf EA PGS diskutiert und der Schluss gezogen, dass es nicht genügend Beweise für eine langfristige Selektion gegen genetische Varianten in Assoziation mit EA gibt.

Einleitung

Die Erforschung menschlichen Verhaltens oblag Jahrzehnte lang fast ausschließlich den Fachgebieten der Soziologie und Psychologie. Obwohl bereits im frühen 19. Jahrhundert durchgeführte Zwillings-Studien Anlass für die Annahme einer gewisse Relevanz genetischer

Komponenten für soziale Eigenschaften nahelegten, fehlten schlichtweg die Methoden für eine ganzheitliche Untersuchung der genetischen Grundlagen des menschlichen Verhaltens. Überdies erwies sich das Verhältnis zu den Disziplinen der Biologie und Genetik aufgrund unterschiedlicher Zugänge, aber auch von Fehlentwicklungen wie wissenschaftlichem Rassismus oder eugenischer Denkweisen als schwierig (Duster, 2006; Freese & Shostak, 2009; Jensen, 1969; Hernstein & Murray, 1996).

Ein Großteil der Resultate aus Kandidatengen-Studien, wie etwa zu externalisierendem und antisozialem Verhalten (Caspi et al., 2002) wurde im Laufe der Jahre aufgrund von mangelnder Replizierbarkeit, problematischer Signifikanznormen, zu geringer Aussagekraft und Publikationsbias revidiert (Duncan, 2014; Hewitt, 2012). Lediglich in Ausnahmefällen, wie dem *APOE*-Gen, welchem eine maßgebliche Rolle für die Entwicklung von Alzheimer nachgewiesen werden konnte, zeigten sich die Resultate als beständig (Adams et al., 2016). In den vergangenen Jahren konnten jedoch aufgrund neuer methodischer Zugänge, effektiverer Genotypisierung, und zunehmender Rechenleistung bahnbrechende Erkenntnisse über die Prädisposition sämtlicher menschlicher Eigenschaften hervorgebracht werden. Diese legen eine unbestreitbare Relevanz für jene Zweige der Sozialwissenschaften die sich mit individuellen Unterschieden im Verhalten und dessen Auswirkungen befassen nahe (Mills & Tropf, 2020). Durch das Aufkommen von GWAS (genome-wide association studies) im Jahr 2005, erschloss sich die Möglichkeit Assoziationen zwischen Phänotypen und „SNPs“ (Einzelbasenunterschiede im Genom) innerhalb von abgegrenzten Populationen herzustellen (Martin et al., 2017). Es handelt sich dabei um hypothesenfreies genetisches Data-mining mit dem Ziel, genetische Loci mit bestimmten Merkmalen in Verbindung zu setzen. Die Identifikation eines Loci als codierend unterliegt hierbei einer Bonferroni-korrigierten Signifikanzschwelle von $p < 5 \times 10^{-8}$ ($p < 0.00000005$), da auf die starke Korrelation zwischen codierenden und nahegelegenen nicht codierenden genetischen Varianten aufgrund von linkage disequilibrium kontrolliert werden muss, um falsch-positive Assoziationen zu vermeiden (Mills & Tropf, 2020). Des Weiteren wird in Metaanalysen unter Kombination verschiedener Datensets die Samplesize maximiert, um insbesondere bei komplexen polygenen Merkmalen die Präzision zu erhöhen und weitere, teils seltene, genetische Varianten zu identifizieren (Dudbridge, 2013). Da bei jenen komplexen polygenen Merkmalen, zu welchen die Verhaltensmerkmale zählen, jede SNP nur für einen sehr geringen Anteil der Merkmalsvarianz in einer Population verantwortlich gemacht

werden kann (Chabris et al., 2013), sind nicht nur große Datensets, sondern statistische Konstrukte wie polygenic scores (PGS) notwendig um die akkumulierte genetische Prädisposition jener Merkmale zu erfassen. Polygenic scores aggregieren additiv multiple genetische Marker (SNPs) wodurch der genetische Einfluss auf einen Phänotypen in einer einzigen quantitativen Variable dargestellt werden kann. Dabei werden die Daten eines Merkmals, stammend aus einer GWAS, auf eine neue Stichprobe angewandt, wobei die genetischen Varianten jeder Person mit der Effektgröße der GWAS gewichtet und summiert werden (Harden & Koellinger, 2020). Aufgrund der steigenden Samplesize und der einhergehenden Entdeckung weiterer co-diegender Loci konnten bereits einige aussagekräftige PGS für Eigenschaften wie Körpergröße, Bodymassindex (BMI), aber auch soziologisch relevante Eigenschaften wie Bildungserfolg (Educational attainment - EA) (Lee et al., 2018), Fruchtbarkeit (Fertility) (Barban et al., 2016; Mathieson et al., 2020; Mills et al., 2020) oder Risikoverhalten (Risky behavior) (Linnér et al., 2019) konstruiert werden. Die verhältnismäßig triviale Handhabung dieser Konstrukte ermöglicht speziell WissenschaftlerInnen außerhalb der Genetik eine reibungslose Integration von genetischen Daten zur Verstärkung statistischer Aussagekraft als Kontrollvariable (Rietveld, 2013), zur Untersuchung von Gen-Umwelt-Interaktionen (Barcellos et al., 2018), oder zur Verminderung von Heterogenität in Untersuchungsgruppen (DiPrete et al., 2018). Mittels einer Methode namens "Linkage disequilibrium score regression", welche aufzeigt wie wahrscheinlich es ist, dass die Anlagen verschiedener Eigenschaften auf den selben genetischen Loci zu finden sind (Pleiotropie¹), werden summary statistics von GWAS verglichen und genetische Korrelationen zwischen den Eigenschaften errechnet (Finucane et al., 2015). Das entsprechende Werkzeug ist online verfügbar (<https://github.com/bulik/ldsc>). Die genetische Grundlage von Educational attainment ist mit über 1300 identifizierten Loci hierbei eine der am weitesten erforschten im Feld (Lee et al., 2018). Mit einem R^2 von ca. 11% kann der polygenic score von EA mit der Effektgröße traditioneller sozialwissenschaftlicher Variablen, wie dem Zusammenhang von Familieneinkommen und Bildungsgrad, abgewogen werden (Harden & Koellinger, 2020). Es gibt Evidenz, dass ein höherer genetischer Score von EA mit dem früheren Erreichen von Entwicklungsmeilensteinen in jungen Jahren, sowie einer schnelleren Entwicklung der Lesekompetenz assoziiert ist (Belsky et al., 2016). Außerdem weisen Menschen die mit mehr EA-Varianten geboren werden eine im Vergleich zu ihren Familien höhere

¹ Pleiotropie bezeichnet in der Genetik die Beeinflussung zweier oder mehr voneinander unabhängiger Merkmale durch ein Gen (Mills & Tropf, 2020).

Aufwärtsmobilität in Bezug auf Einkommen, beruflichen Status und Bildungsniveau auf (Belsky et al., 2018). Eine Publikation von Mills & Tropf (2020) konnte überdies für EA starke genetische Korrelationen mit Eigenschaften wie dem Alter bei der ersten Geburt (+), der Anzahl an geborenen Kindern (-), dem Alter bei Beginn des Rauchens (+) und weiteren Eigenschaften, mit Vorbehalt auf umwelt-vermittelte Einflüsse sowie Pleiotropie, feststellen.

Der Zusammenhang von Bildung und Kinderzahl ist für ForscherInnen nicht erst seit dem Aufkommen von GWASs von Interesse. Vorhergegangene Studien zum Einfluss verschiedener Phänotypen auf den reproduktiven Erfolg konnten Anzeichen für eine fortlaufende natürliche Selektion in kontemporären westlichen Gesellschaften feststellen. Das Hauptaugenmerk der Studien liegt neben dem negativen Einfluss von Bildung (Huber et al., 2010; Weeden et al., 2006), auf Faktoren wie sozioökonomischem Status (Fieder & Huber, 2007; Skirbekk, 2008; Hopcroft, 2006) und Geschlecht (Fieder & Huber, 2007; Hopcroft, 2006). Dennoch kann nur von natürlicher Selektion gesprochen werden, wenn die selektierten Phänotypen auch mit kovariierenden Genotypen assoziiert sind (Beauchamp, 2016). Dementsprechend unterliegen die Ergebnisse der rein phänotypischen Erforschung diverser Annahmen über die Beziehungen zwischen Geno- und Phänotypen sowie der Annahme, dass alle korrelierten Phänotypen mit kausalem Einfluss auf den reproduktiven Erfolg in natürlichen Populationen berücksichtigt werden können (Lande & Arnold, 1983; Morrissey et al., 2010).

Die in den vergangenen Jahren aufgekommenen molekulargenetischen Methoden ermöglichen letztlich eine Erweiterung der phänotypischen Forschung durch den Einbezug genetischer Komponenten. Eine Auswahl an Zwillings- und Familienstudien konnte bereits eine genetische Prädisposition für die Gesamtzahl an Kindern (Number of children ever born -NEB) und das Alter bei der Geburt des ersten Kindes (Age at first birth - AFB) feststellen (Briley et al., 2017; Tropf et al., 2015; Zietsch et al., 2014). Weiterführend konnte eine 2016 durchgeführte GWAS mit über 345.000 ProbandInnen, 12 Loci in Zusammenhang mit dem Alter bei der ersten Geburt (Age at first birth - AFB) und der Gesamtzahl an gezeugten Kindern (Number of children ever born - NEB) identifizieren (Barban et al., 2016). In einer Erweiterungsstudie aus dem Jahr 2020 konnten weitere 43 Loci in Assoziation mit reproduktions-biologisch relevanten Aspekten lokalisiert werden. Weiters wurden Indizien für eine fortlaufende natürliche

Selektion auf NEB-erhöhende Allele, mittels der Integration von aDNA-Daten (ancient DNA), gesammelt (Mathieson et al., 2020).

Unter Einbezug der Bildungskomponente lieferten Studien jedoch Anzeichen für eine Selektion gegen erhöhten Bildungserfolg, auch auf genetischer Ebene. Eine Studie von Beauchamp (2016) konnte anhand phänotypischer sowie genetischer Daten durch den PGS EA ($n = 11.850$) bereits einen signifikant negativen Einfluss von Bildungserfolg auf die Kinderzahl feststellen. Auch eine Studie aus Island ($n = 129.808$) kam zu dem Schluss, dass der PGS EA mit verzögerter Reproduktion und weniger geborenen Kindern assoziiert werden kann (Kong et al., 2016). Untersuchungen eines kausalen Zusammenhangs zwischen Bildung und dem Alter bei der ersten Geburt lieferten in Zwillings- und Familienstudien gemischte Resultate, wobei das familiäre Umfeld und nicht genetische Faktoren als die treibende Kraft identifiziert wurde (Tropf & Mandemakers, 2017; Rodgers et al., 2008).

In der vorliegenden Studie soll dieser bildungs-medierte Einfluss auf die Kinderzahl durch weitere phänotypische und genetische Komponenten spezifiziert und ergänzt werden. Ziel ist es einen vertieften Einblick in die prädiktive Qualität der aktuellen PGSs von EA und NEB unter Einfluss weiterer phänotypischer Variablen wie Geschlecht, Einkommen und realen Ausbildungsjahren zu erlangen. Dabei soll die relative Effektstärke der Prädiktorvariablen auf die Kinderzahl mittels multipler linearer Regression evaluiert werden.

Da PGS keineswegs unabhängig von unbeobachteten umwelt-gesteuerten Einflüssen wie Erziehung und Nachbarschafts-charakteristika sind und genetische Einflüsse oft durch die Umwelt bedingt und vermittelt werden, ist es notwendig innerhalb der gleichen Population multivariat auf Geschlecht, sozioökonomischem Status, Alterskohorte sowie Land zu korrigieren (Mostafavi et al., 2020; Tropf et al., 2017). Speziell Geschlechterunterschiede sind für Konstrukte wie PGSs bisher weitgehend unerforscht, wobei eine Studie von Khramtsova et al. (2019) auf gravierende Unterschiede, speziell bei komplexen Eigenschaften, aufmerksam machen konnte. Dementsprechend wird in dieser Studie der Fokus auf geschlechtsspezifische Unterschiede und überdies auf mögliche sozioökonomische Einflüsse, gemessen an Einkommensvariablen gesetzt.

Methoden

In der vorliegenden Analyse wird der Einfluss mehrerer unabhängiger Prädiktorvariablen auf die Anzahl an geborenen Kindern mittels multipler linearer Regressionen am Datensample der HRS (Health and Retirement Survey) untersucht. Dabei werden sowohl genetische Komponenten in Form der PGSs von Educational attainment und Number of children ever born, als auch phänotypische Komponenten wie die tatsächlichen Ausbildungsjahre, Geschlecht und zwei unterschiedliche Messungen des Einkommens in die Modelle integriert und auf ihre relative Effektstärke und Signifikanz überprüft. Um verzerrte Effekte zu minimieren, wurde auf das Alter sowie auf Confounding durch Populationsunterschiede von Personen europäischer Abstammung kontrolliert. Dies geschah durch die Integration von zehn "ancestry-specific principal components" (PCs) in das Modell. Die PCs korrigieren für jene genetischen Aspekte der gemeinsamen Abstammung, die fälschlicherweise mit dem PGS und dem jeweiligen Ergebnis korrelieren und dadurch zu falsch-positiven bzw. falsch-negativen Resultaten führen könnten (Price et al., 2006).

Datensample

Diese Studie bezieht sich auf die Daten der Health and Retirement Study (HRS) durchgeführt von der University of Michigan. Die Studie gibt einen Überblick über körperliche und geistige Gesundheit, finanziellen Status, Versicherungsschutz, Arbeitsmarktstatus und Ruhestandsplanung einer alternden USA. Seit 1992 nehmen mehr als 22.000 US-AmerikanerInnen europäischer Abstammung alle zwei Jahre an der Befragung teil. Im Zeitraum von 1992 und 2010 konnte das ursprüngliche Sample auf insgesamt sieben Subsamples verschiedener Alterskohorten aufgestockt werden. Alle befragten Personen wurden zwischen 1900 und 1992 geboren, wobei die primären Befragten ein Alter von über 50 Jahren aufweisen und die EhepartnerInnen der Befragten unabhängig vom Alter interviewt wurden (Servais, 2010). In der vorliegenden Studie wird nur auf die Daten der primären ProbandInnen zugegriffen, wobei der reproduktive Erfolg im Alter von 50 Jahren als abgeschlossen angenommen werden kann. Überdies wird das Alter als Kontrollvariable in die multiple lineare Regression integriert, um unentdeckte Einflüsse zu eliminieren. Um Effekte der genetischen Populationsschichtung (population stratification) entgegenzuwirken wurde die Analyse auf nicht verwandte Personen europäischer Abstammung reduziert (n=12.090) (HRS Documentation Report, 2018). Von

einer Teilstichprobe der HRS-Teilnehmer wurde zwischen den Jahren 2006 und 2008 DNA-Proben gesammelt, wobei 95% der genotypisierten Probanden 1953 oder früher geboren wurden. Um sicher zu gehen, dass die Ergebnisse der genotypisierten Stichprobe mit den Ergebnissen der nicht genotypisierten Probanden übereinstimmen, wurden nur Personen einbezogen, die bis zum Jahr 2008 nach der Zahl der geborenen Kinder gefragt wurden. Dadurch verblieben 6.414 weibliche und 5.436 männliche Probanden mit phänotypischen Daten und 3.416 nicht verwandte Frauen, sowie 2.571 nicht verwandte Männer die einer Genotypisierung unterzogen worden sind (Beauchamp, 2016).

Konstruktion der Polygenic Scores und Variablen

PGS Educational attainment (EA)

Der PGS für den Bildungserfolg basiert auf den Resultaten der 2018 durchgeführten Studie des Social Science Genetic Association Consortium (SSGAC). Die Konstruktion basiert auf der Metaanalyse der Untersuchungs- und Replikationsstichprobe, welche 405.073 ProbandInnen umfasst, zuzüglich einer weiteren Replikationsstudie mit 726.808 ProbandInnen (insgesamt 1.131.881 Individuen). Eine dritte Replikationsanalyse wurde mit den Daten von 23andMe (n=368,890) durchgeführt welche ebenfalls in den verwendeten Score von EA integriert wurden. Aus über 10.2 Millionen SNPs konnten genomweit 1.271 Loci signifikanter SNPs identifiziert werden, wobei in allen Kohorten imputierte SNPs aus dem 1000 Genomes Referenzpanel (1000G) integriert wurden. Aufgrund von Datennutzungsvereinbarungen wurden die ursprünglich in der GWAS integrierten HRS- und 23andMe-Daten entfernt, weshalb die von der SSGAC zur Verfügung gestellten SNP-Gewichte nicht zu 100% mit der HRS-Kohorte übereinstimmen. Die studienspezifischen GWASs kontrollierten für die ersten zehn Hauptkomponenten der genetischen Daten, das Alter, einen Indikator für weibliches Geschlecht, Interaktionen zwischen Alter und weiblichem Geschlecht, weiteren studienspezifischen Variablen und Großereignisse wie politische Veränderungen oder Kriege die einen möglichen Einfluss auf den Bildungszugang gehabt haben können, mittels Dummy-Variablen. Die PGSs europäischer Abstammung umfassen 1.274.056 SNPs, die sich zwischen der genetischen Datenbank von HRS und der GWAS-Metaanalyse überlappen. Anschließend wurden die PGSs abhängig von der Ethnizität standardisiert und normalverteilt (Mittelwert = 0, Standardabweichung = 1). Die phänotypische Messung von EA erfolgte über sieben international vergleichbare Kategorien

von Bildungsniveau, welche auf US-Schuljahre transformiert wurden (HRS Documentation Report, 2018; Lee et al., 2018). In dieser Arbeit wird ausschließlich der beschriebene PGS von Educational attainment in Kombination mit den 23andMe Daten für die Analysen herangezogen.

PGS Number of children ever born (NEB)

Der PGS für die Gesamtanzahl an gezeugten Kindern basiert auf den Ergebnissen einer Studie des *Sociogenome consortium*, durchgeführt im Jahr 2016. In der Metaanalyse wurden die Daten von 243.072 Männer und Frauen aus 62 Kohorten europäischer Abstammung erfasst. Dabei absolvierten 2,4 Millionen SNPs, imputiert durch NCBI Build 37 HapMap Phase 2 Daten, die Qualitätskontrolle positiv. Die Kontrolle der identifizierten Assoziationen erfolgte auf die zehn Hauptkomponenten der genetischen Daten sowie auf das Geburtsjahr der ProbandInnen um Confounding durch Populationsunterschiede und Geburtskohorteneffekte zu minimieren. Die NEB wurde nur bei jenen ProbandInnen ausgewertet, welche bereits die reproduktive Phase abgeschlossen haben (≥ 45 Jahre für Frauen und ≥ 55 Jahre für Männer). Aus den Ergebnissen der Analyse wurden PGS getrennt nach Geschlecht aber auch als Gesamtscore errechnet und zur Verfügung gestellt. Genomweit konnten drei Loci signifikant mit NEB assoziiert werden, zwei davon mit beiden Geschlechtern und einer nur bei Männern. ($n = 103,736$) Der Gesamt-PGS für europäische Abstammung umfasst 764.577 SNPs, die sich zwischen der genetischen Datenbank von HRS und der GWAS-Metaanalyse überschneiden. Davon überlappen sich 763.060 SNPs bei der weiblichen Stichprobe und 763.620 SNPs bei der männlichen Stichprobe. Die drei PGSs sind abhängig von der Ethnizität standardisiert und normalverteilt (Mittelwert = 0, Standardabweichung = 1) (HRS Documentation Report, 2018; Barban et al., 2016). Für die in dieser Arbeit vorgenommenen Analyse wurden ausschließlich die geschlechtsspezifischen Variablen für Männer und Frauen, nicht jedoch der Gesamtscore verwendet (HRS Documentation Report, 2018).

Ausbildungsjahre

Die Variable der Ausbildungsjahre (years of education) wird in der HRS durch die Betrachtung der Berichte aller Datenwellen zugeordnet. Dabei wird die erste nicht fehlende Angabe aus allen Wellen verwendet. Sollten alle Angaben fehlen wurde auf die Angabe der Jahre der Schulbildung (years of schooling) zurückgegriffen (HRS Documentation Report, 2018).

Einkommensvariablen

Repräsentativ für die Studie werden die Einkommensdaten aus der ersten Erhebungswelle der HRS herangezogen. Da zum Erhebungszeitpunkt der weiteren Wellen keine erwähnenswerten Wirtschaftseinbrüche festgestellt werden konnten und grundsätzlich mit einem linearen Anstieg des Einkommens zu rechnen ist wurde lediglich auf altersbedingte Einflüsse durch das Geburtsjahr geprüft. Aus Kontrollzwecken werden die folgenden zwei unterschiedlichen Erfassungsmöglichkeiten des Einkommens in separaten Regressionen verwendet: Gesamteinkommen (Earnings) und Rentenversicherungsvermögen (Social-Security Wealth). Das Gesamteinkommen setzt sich aus einer Kombination verschiedener Komponenten zusammen und wurde mittels Selbstbericht im Fragebogen durch Angabe in nominalen Dollar erhoben. Die Gesamteinkommensvariable (R1IEARN: Variablenkurzbezeichnung aus dem HRS) ergibt sich aus der Summe des Lohnes/Gehalts, weiterer Boni/Überstundenvergütungen/Provisionen/Trinkgelder, Einkommen aus Zweitjobs, Einkünfte aus der US Military Reserve und weiterer gewerblicher Einkünfte der Befragten. Das Rentenversicherungsvermögen (R1SSWRER: Variablenkurzbezeichnung aus dem HRS) ergibt sich in den USA durch eine monatliche prozentuale Abgabe des Einkommens von Angestellten und Selbstständigen. In der HRS wird das voraussichtliche Rentenversicherungsvermögen auf Grundlage der offiziellen Einkommensaufzeichnungen der Social Security Administration (SSA), unter Annahme eines Anspruchsalters von 62 Jahren berechnet. Bei Individuen, die das Anspruchsalter noch nicht erreicht haben wurden Prognosen für die zukünftigen Einnahmen bis zum Alter von 62 Jahren erstellt und integriert. Weiterführende Informationen über die Konstruktion der Variable können der „Prospective Social Security Wealth Measures of Pre-Retirees documentation“ der HRS entnommen werden (HRS Documentation Report, 2018).

Um möglichen Differenzen zwischen der selbstberichteten, sämtliche Einkünfte umfassenden, Einkommensvariable und den real versteuerten offiziellen Einkünften gemessen am Rentenversicherungsvermögen zu identifizieren werden beide in separaten Regressionen einbezogen. Zur Überprüfung der Überschneidung der beiden Variablen wurde eine Pearson-Korrelation durchgeführt wobei ein Korrelationskoeffizient von $r = 0,467$ ($p \leq 0,01$) deutliche Differenzen nachweist.

Ergebnisse

Um die Erklärungsbeiträge der einzelnen Variablen auf die Kinderzahl einschätzen zu können wird die Effektstärke der Variablen in standardisierten Koeffizienten angegeben. Aufgrund der Stichprobengröße und der genau definierten Rahmenbedingungen des Samples ist nicht mit irreführendem Samplebias zu rechnen. **Tabelle 1** stellt die Ergebnisse der Regressionsanalyse unter Integration der Einkommensvariable „Gesamteinkommen“ dar. Der PGS von NEB weist sowohl für Frauen ($\beta = 0,394$; $p = <0,0001$) als auch für Männer ($\beta = 0,492$; $p = <0,0001$) den höchsten prädiktiven Effekt für die Kinderzahl auf. Der PGS EA hingegen verliert unter Einfluss der anderen Variablen den signifikant negativen Effekt auf die Kinderzahl sowohl bei Frauen ($\beta = 0,026$; $p = 0,145$) als auch bei Männern ($\beta = 0,010$; $p = 0,629$). Zur Referenz zeigt **Tabelle 3** den hochsignifikant negativen Zusammenhang von EA auf die Gesamtzahl an Kindern in einer separaten Regression unter Kontrolle auf Alter und "ancestry-specific principal components", wie er auch in den vorhergehenden Publikationen von Beauchamp (2016) und Kong et al. (2016) festgestellt werden konnte. Die Zahl der Ausbildungsjahre hat bei Frauen einen bedeutenden, hochsignifikant negativen Effekt auf die Kinderzahl ($\beta = -0,157$; $p = <0,0001$). Bei den männlichen Probanden kann ebenfalls ein etwas abgeschwächter jedoch immer noch signifikanter negativer Einfluss festgestellt werden ($\beta = -0,052$; $p = 0,011$). Beim Gesamteinkommen von Frauen kann ein hochsignifikant negativer Zusammenhang mit der Zahl an geborenen Kindern unter relativ geringer Effektstärke nachgewiesen werden ($\beta = -0,067$; $p = 0,0002$). Dieser Effekt dreht sich bei Männern um, wonach ein sehr signifikanter positiver Effekt auf die Kinderzahl gemessen wurde ($\beta = 0,051$; $p = 0,009$).

In **Tabelle 2** wurde die Variable Gesamteinkommen durch das Rentenversicherungsvermögen ersetzt um potenzielle Unterschiede zwischen den selbstberichteten Einkommensdaten inklusive sämtlicher möglicher Zulagen und dem Einkommen gemessen an offiziellen Steueranteilen festzustellen. Effektstärke und Signifikanz des PGS NEB verringern sich sowohl bei Männern ($\beta = 0,485$; $p = <0,0001$) als auch bei Frauen ($\beta = 0,373$; $p = <0,0001$) in minimalem Ausmaß. Vergleichbare minimale Veränderungen sind auch bei den Variablen PGS EA und der Zahl der Ausbildungsjahre für beide Geschlechter zu beobachten, wobei keine Änderung der Signifikanzniveaus festzustellen sind. Der Effekt des Verdienstes gemessen am Rentenversicherungsvermögen von Männern ist unter geringer Effektstärke signifikant positiv ($\beta = 0,050$; $p = 0,016$) und unterscheidet sich somit kaum von der in **Tabelle 1** dargestellten

Gesamteinkommensvariable. Bei den weiblichen Probandinnen konnte ebenfalls ein hochsignifikant negativer Effekt auf die Kinderzahl festgestellt werden, wobei ein merklich höherer Erklärungsbeitrag für die Variable Rentenversicherungsvermögen ($\beta = -0,169$; $p = <0,0001$) als für das Gesamteinkommen ($\beta = -0,067$; $p = 0,0002$) festgestellt wurde.

Tabelle 1: *Multiple lineare Regression der $\sqrt{\text{Kinderzahl}}$ regressiert auf PGS Educational attainment, PGS Number of children ever born, die Anzahl der Ausbildungsjahre und Gesamteinkommen, getrennt nach Geschlecht. Standardisierte Koeffizienten und Signifikanzniveau der Ergebnisse sind angegeben.*

Variable	Frauen		Männer	
	Standardisierte Koeffizienten (Beta-Wert)	Sig. p	Standardisierte Koeffizienten (Beta-Wert)	Sig. p
PGS Number of children ever born (NEB)	0,394	< 0,0001	0,492	< 0,0001
PGS Educational attainment (EA)	0,026	0,145	0,010	0,629
Zahl der Ausbildungsjahre	-0,157	< 0,0001	-0,052	0,011
Gesamteinkommen	-0,067	0,0002	0,051	0,009

Anmerkung: Die überprüfte Kontrollvariable Alter sowie die zehn "ancestry-specific principal components" (PCs) zur Vermeidung von unbeobachtetem Confounding durch Populationsunterschiede werden nicht in den Ergebnissen angeführt.

Tabelle 2: *Multiple lineare Regression der $\sqrt{\text{Kinderzahl}}$ regressiert auf PGS Educational attainment, PGS Number of children ever born, die Anzahl der Ausbildungsjahre und Rentenversicherungsvermögen, getrennt nach Geschlecht. Standardisierte Koeffizienten und Signifikanzniveau der Ergebnisse sind angegeben.*

Variable	Frauen		Männer	
	Standardisierte Koeffizienten (Beta-Wert)	Sig. p	Standardisierte Koeffizienten (Beta-Wert)	Sig. p
PGS Number of children ever born (NEB)	0,373	< 0,0001	0,485	< 0,0001
PGS Educational attainment (EA)	0,029	0,124	0,007	0,753
Zahl der Ausbildungsjahre	-0,138	< 0,0001	-0,048	0,023
Rentenversicherungsvermögen	-0,169	< 0,0001	0,050	0,016

Anmerkung: Die überprüfte Kontrollvariable Alter sowie die zehn "ancestry-specific principal components" (PCs) zur Vermeidung von unbeobachtetem Confounding durch Populationsunterschiede werden nicht in den Ergebnissen angeführt.

Tabelle 3: Lineare Regression der $\sqrt{\text{Kinderzahl}}$ regressiert auf PGS Educational attainment getrennt nach Geschlecht. Standardisierte Koeffizienten und Signifikanzniveau der Ergebnisse sind angegeben.

Variable	Frauen		Männer	
	Standardisierte Koeffizienten (Beta-Wert)	Sig. p	Standardisierte Koeffizienten (Beta-Wert)	Sig. p
PGS Educational attainment (EA)	-0,054	< 0,0001	-0,033	0,016

Anmerkung: Die überprüfte Kontrollvariable Alter sowie die zehn "ancestry-specific principal components" (PCs) zur Vermeidung von unbeobachtetem Confounding durch Populationsunterschiede werden nicht in den Ergebnissen angeführt.

Diskussion

Die Ergebnisse dieser explorativen Forschungsarbeit konnten im vorliegenden HRS-Sample einen starken Zusammenhang der Kinderzahl mit dem Einkommen und den absolvierten Bildungsjahren aufzeigen. Für die beiden untersuchten PGS von Educational attainment und Number of children ever born wurden unterschiedliche Abänderungen der Einflussstärke unter Einbezug der phänotypischen Variablen verzeichnet. Der genetische Score von NEB weist den stärksten relativen Einfluss auf die Kinderzahl auf und verdeutlicht somit die unverkennbare Bedeutung der additiven genetischen Prädisposition für den reproduktiven Erfolg. Vergleicht man die Werte mit jenen einer einfachen linearen Regression der Kinderzahl auf NEB (Anhang: Tabelle 4), so kann bis auf eine leichte Erhöhung der Effektstärke kaum eine Änderung der standardisierten Koeffizienten festgestellt werden. Anzumerken ist, dass der PGS NEB für Männer eine noch höhere prädiktive Effektstärke aufweist ($\beta +0,1$) als für Frauen. Dies könnte darauf zurückgeführt werden, dass bei der Erstellung der PGSs von NEB zwei genomweit signifikante Loci für beide Geschlechter und eine weiterer ausschließlich für Männer identifiziert werden konnte (Barban et al., 2016). Folgestudien und die einhergehende Erweiterung der Samplesize sind erforderlich um für eine weitere Präzisierung der Scores beider Geschlechter zu sorgen und den Unterschied eventuell einzuebnen.

Die in bisheriger Literatur beobachtete Selektion gegen mit Bildungserfolg assoziierte genetische Varianten (Beauchamp, 2016; Kong et al., 2016) kann in der vorliegenden Studie nur in der einfachen linearen Regression auf den PGS EA (Tabelle 3) nachgewiesen werden. In der multivariaten Regression geht der signifikante Zusammenhang bei beiden Geschlechtern

verloren. Ebenso ist dies der Fall unter Einbezug beider Einkommensvariablen (Tabelle 1 & Tabelle 2). Die Integration des PGS NEB, des Einkommens, sowie der realen Bildungsjahre scheint den Einfluss des PGS EA als Prädiktor zu mitigieren. Dies könnte dadurch zu begründen sein, dass der PGS EA trotz der fundierten Konstruktion mittels großem Datensample (Lee et al., 2018) weiterhin nur ein unvollkommener Proxy für den tatsächlichen genetischen Score darstellt, wodurch die statistische Aussagekraft der Analyse eingeschränkt wird. Zudem ist es möglich, dass der Effekt schlichtweg von den zusätzlichen Variablen, wie dem hoch signifikant negativen Einfluss der tatsächlichen Ausbildungsjahre, überlagert wird und daher seine Signifikanz verliert. Erwähnenswert sind die im Anhang angeführten Ergebnisse, welche mittels multipler Regression unter Einbezug des PGS EA ohne die zusätzlichen Daten von 23andMe festgestellt werden konnten (Tabelle 4 & Tabelle 5). Der Effekt dieses EA PGS war für Frauen unter Einbezug der Einkommensvariable ‚Rentenversicherungsvermögen‘, konträr zu bereits erwähnter Literatur, signifikant positiv auf die Kinderzahl ($\beta = 0,044$; $p = 0,023$). Dementsprechend würden Frauen mit höherem genetischem Score für Bildung signifikant mehr Kinder zeugen. Einige Studien konnten in der Vergangenheit, passend zu diesen Ergebnissen, aufweisen, dass sich der im 20. Jahrhundert beobachtete Trend des negativen Zusammenhangs von Fertilität und wirtschaftlicher und sozialer Entwicklung, in den letzten Jahren unter der höchstentwickelten Bevölkerungsschicht vermehrt umkehrt (Myrskylä et al., 2009). Bei Frauen in den USA wurde indes der Zusammenhang von hoher Bildung und gesteigerter Kinderzahl auf die Substitution von Kindererziehung und Haushaltsführung durch leistbare Marktdienstleistungen zurückgeführt (Hazan & Zoabi, 2015). In Anbetracht des signifikant negativen Einflusses der tatsächlichen Bildungsjahre auf die Kinderzahl, sowie bisheriger Ergebnisse von Untersuchungen desselben Datensamples (Beauchamp, 2016) können die Werte des PGS EA ohne 23andMe jedoch nicht nachvollziehbar diesem Phänomen untergeordnet werden und bedürfen daher weiterer Klärung. Aus Gründen der Nachvollziehbarkeit wurde deshalb der polygenic score von EA in Erweiterung der Daten von 23andMe gewählt.

Bei Betrachtung von Ausbildungsjahren und Einkommen können die Ergebnisse bisheriger phänotypischer Studien untermauert werden (Huber et al., 2010; Hopcroft, 2006; Skirbekk, 2008). Die Zahl der tatsächlichen Ausbildungsjahre hat im Vergleich zum PGS EA bei beiden Geschlechtern einen signifikant negativen Einfluss auf die Kinderzahl wobei die Effektstärke bei Frauen etwas bedeutender ist. Sowohl für die Variablen ‚Gesamteinkommen‘ als auch für

die Variable ‚Rentenversicherungsvermögen‘ konnte ein signifikant negativer Einfluss auf die Anzahl der Nachkommen bei Frauen festgestellt werden. Bei Männern jedoch war dieser Einfluss signifikant positiv. Dieses Ergebnis bestätigt die Beobachtungen vorhergehender Studien (Hopcroft, 2006). Interessanterweise ergibt sich bei Frauen eine erwähnenswert höhere Effektstärke unter Einbezug des Rentenversicherungsvermögens ($\beta = -0,169$) als beim Gesamteinkommen ($\beta = -0,067$). Bei den männlichen Probanden bleibt der Wert jedoch nahezu unverändert. Bezüglich der Effektstärke liegt der negative Einfluss von Ausbildung und Einkommen (ermittelt durch die offizielle Einkommensbesteuerung) bei Frauen somit auf einem sehr ähnlichen Niveau. Womöglich wird der negative Effekt des Einkommens auf die Kinderzahl bei Frauen durch die Addition von un versteuerten Nebeneinkünften oder abhängig von Eigenauskunft tendenziell etwas abgefedert. Zusätzliche im Gesamteinkommen integrierte Einkünfte wie Trinkgeld könnten beispielsweise zu mehr Einkommen bei gleicher Arbeitszeit führen und somit den Effekt nachvollziehbar abschwächen.

Schaut man sich das Verhältnis von Bildung zu Einkommen an so kann besonders in westlichen kapitalistischen Gesellschaften ein eindeutiger sozialer Zusammenhang zwischen dem Bildungsgrad und der Höhe des Einkommens festgestellt werden (Hoeller et al., 2014; Domanski, 2005). Da der höchste Bildungsgrad statistisch stark von der Intelligenz beeinflusst wird (Sewell & Shah, 1967; Deary et al., 2007) und bereits ein bemerkenswerter genetischer Zusammenhang zwischen den Veranlagungen für Intelligenz und Einkommen in einer britischen GWAS ermittelt werden konnte (Hill et al., 2019), ist auch ein möglicher Einfluss der genetischen Prädisposition für Bildungserfolg auf das Einkommen von Interesse. In der Studie von Hill et al. konnten mittels genetischer Korrelationsanalysen bereits auffällige Überlappungen von Educational attainment (Lee et al., 2018) und Einkommen mit einer Vielzahl an weiteren Merkmalen festgestellt werden. Zur weiteren Ergründung dieser Beziehung wurde in der hier verwendeten HRS-Datensample ein möglicher Effekt des PGS Educational attainment (mit den Daten von 23andMe) auf das Gesamteinkommen mittels Regressionsanalyse errechnet (Anhang: Tabelle 7), wobei ein hoch signifikanter genetischer Einfluss auf das Einkommen gefunden werden konnte ($\beta = 0,105$; $p = <0,0001$). Dementsprechend gibt es Anzeichen, dass neben umweltvermittelten Einflüssen auch die genetischen Prädispositionen für Intelligenz und Bildungserfolg Einfluss auf sozioökonomische Ungleichheiten in der heutigen Gesellschaft

nehmen. Weiterführende Forschung ist für eine akkurate Klärung der genetischen Verhältnisse, sowie der Effektstärke jedoch von Nöten.

Konkludierend können die vorliegenden Ergebnisse keinen Beweis für eine fortlaufende Selektion gegen Bildung auf genetischer Ebene (Beauchamp, 2016), sondern lediglich auf phänotypischer Ebene liefern. Fest steht, dass die bisher beobachteten Selektionsindizien mit höchster Wahrscheinlichkeit zu einem bedeutenden Anteil auf Gen-Umwelt-Interaktionen zurückzuführen ist (Mills & Tropf, 2020; Rimfeld et al., 2018; Harden et al., 2020), wodurch sich die Resultate nicht barrierefrei auf andere sozioökonomische Schichten und Populationen übertragen lassen. Kennzeichnenderweise konnten in sozioökonomisch stark benachteiligten Ländern positive Effekte von Bildung auf die Fertilität am unteren Ende der Bildungsskala festgestellt werden (Castro, 1995). Eine bahnbrechende Arbeit von Turkheimer et al. konnte bereits im Jahr 2003 zeigen, dass die Heritabilität von IQ am höchsten bei Familien mit hohem sozioökonomischem Status und am niedrigsten bei Familien mit niedrigem sozioökonomischem Status ist. Diesbezüglich ist dem bioökologischen Modell entsprechend, nicht zu missachten, dass genetische Einflüsse in stabilen, adaptiven und sozioökonomisch fortgeschrittenen Umgebungen maximiert werden, wodurch eine Realisierung des genetischen Potentials erst ermöglicht wird (Bronfenbrenner & Ceci, 1994). Das verhaltensgenetische Forschungsfeld steht dementsprechend vor zahlreichen offenen Fragen bezüglich der Einbindung umweltbasierter Faktoren aufgrund von informellen Messmodellen, komplexen wechselseitigen Kausalprozessen zwischen Umwelt und Outcome und weiteren Inkonsistenzen. Eine genauere Darstellung der bestehenden Probleme und möglicher Lösungsansätze im Umgang mit Gen-Umwelt-Interaktionen wird im Review-paper von Mills und Tropf (2020) ausgeführt.

Die oben erläuterten gravierenden Unterschiede zwischen den verschiedenen polygenic scores von Educational attainment weisen auf einige Inkonsistenzen in der Bildung und Anwendbarkeit der jeweiligen Scores hin, welche zukünftiger Klärung und Präzisierung bedürfen. Selbst wenn bisherige Studien von Beauchamp (2016) und Kong et al. (2016) eine negative genetische Korrelation von Bildungsniveau und Geburtenzahl finden konnten ist es fraglich ob es sich, in Anbetracht evolutionärer Zeiträume, um Momentaufnahmen oder Jahrhunderte anhaltende Trends mit tiefgreifenden Auswirkungen handelt. Letztlich besteht die Möglichkeit, dass es sich lediglich um ein temporäres transitorisches Phänomen handelt (Fisher, 1958;

Cavalli-Sforza & Bodmer, 1999). Des Weiteren lassen sich die, fast ausschließlich an Personen europäischer Abstammung ermittelten Ergebnisse (Mills & Rahal, 2019; Tropf et al., 2017) weder in ihrer Auswirkungsrichtung noch im Ausmaß auf andere Populationen und Bevölkerungsschichten übertragen, weshalb eine zukünftige Sample-erweiterung auf Personen asiatischer, afrikanischer, afrokaribischer, hispanischer und lateinamerikanischer Abstammung (ancestry) für die Forschung unumgänglich ist (Popejoy & Fullerton, 2016; Mills & Rahal 2020). Eine weitere Studie machte außerdem auf die Abhängigkeit der Vorhersagegenauigkeit der PGS vom Studiendesign, sowie von der Alters- und Geschlechtszusammensetzung der untersuchten GWAS-Stichprobe aufmerksam (Mostafavi et al., 2020). Dieses Problem könnte mit der Entwicklung von Gewichtungsmethoden, welche eine Übertragbarkeit von einer Stichprobe auf andere Zielpopulationen gewährleisten, gelöst werden (Domingue et al., 2017). Die Erhebung von repräsentativeren und vielfältigeren Daten, sowie eine Diversifizierung der Vergleichsgruppen würde jedoch eine langfristige Lösung darstellen (Mills & Tropf, 2020).

Das Aufstreben des Feldes der Verhaltensgenetik und deren Implikationen für die Soziologie werfen auch einige ethische Fragen auf. Es muss gesagt sein, dass es sich in diesem Forschungszweig bei den entdeckten kausalen Zusammenhängen grundsätzlich eher um probabilistische, als deterministische Aussagen handelt, welche keinen Anspruch auf Universalität erheben (Harden & Koellinger, 2020). Dass menschliches Verhalten durch zwischenmenschliche genetische Unterschiede beeinflusst wird, soll keineswegs als Bio-Determinismus fehlinterpretiert werden, da die Ausprägung genetischer Einflüsse maßgeblich von der Umwelt beeinflusst wird und dadurch äußerst variabel ist (Jencks, 1980). Beispielhaft konnten fünf Langzeitstudien auf drei verschiedenen Kontinenten zeigen, dass Kinder aus reichen Familien mit hohem Educational attainment PGS, immer noch mehr sozioökonomischen Erfolg aufweisen als Kinder aus armen Familien mit hohem PGS (Belsky, 2018).

Insbesondere die genetische Erforschung von Bildungserfolg stieß bei einigen Soziologen auf Ablehnung, mit der Begründung, dass es sich um die Hintertür zur Eugenik handle (Bliss, 2018; Duster, 2006). Trotz einiger überzogener Behauptungen (Freese, 2018) konnten einige Arbeiten dennoch wichtige Fragen über die Interpretation und Verwendung von verhaltensgenetischen Forschungsergebnissen aufbringen (Nelson, 2016, Reardon, 2017). Neben denkbaren missbräuchlichen Anwendungsmöglichkeiten, wie beispielweise der Schaffung eines

selektierenden Bildungszuganges abhängig von genetischer Prädisposition, gibt es selbstverständlich eine Vielzahl an nützlichen Anwendungen für PGS. So meint Robert Plomin (2019; Plomin & von Stumm, 2018), dass es neben dem medizinischen Nutzen in Form von präzisiert individualisierten Analyse- und Behandlungsmöglichkeiten auch vorstellbar ist, passgenaue Lerntools für eine maßgeschneiderte Bildungsförderung mittels des EA PGS zu entwickeln. Wie jedoch an den Ergebnissen der hier durchgeführten Studie und einer weiteren Arbeit von Morris et al. (2019) gezeigt wird, ist weiterführende Forschung notwendig, um Vorhersagen durch den EA PGS speziell auf individueller Ebene treffen zu können.

Literaturverzeichnis

- Barban, N., Jansen, R., De Vlaming, R., Vaez, A., Mandemakers, J. J., Tropf, F. C., ... & Hyppönen, E. (2016). *Genome-wide analysis identifies 12 loci influencing human reproductive behavior*. *Nature genetics*, 48(12), 1462-1472.
- Barcellos, S. H., Carvalho, L. S., & Turley, P. (2018). *Education can reduce health differences related to genetic risk of obesity*. *Proceedings of the National Academy of Sciences*, 115(42), E9765-E9772.
- Beauchamp, J. P. (2016). *Genetic evidence for natural selection in humans in the contemporary United States*. *Proceedings of the National Academy of Sciences - PNAS*, 113(28), 7774-7779.
- Belsky, D. W., Domingue, B. W., Wedow, R., Arseneault, L., Boardman, J. D., Caspi, A., ... & Harris, K. M. (2018). *Genetic analysis of social-class mobility in five longitudinal studies*. *Proceedings of the National Academy of Sciences*, 115(31), E7275-E7284.
- Belsky, D. W., Moffitt, T. E., Corcoran, D. L., Domingue, B., Harrington, H., Hogan, S., ... & Caspi, A. (2016). *The genetics of success: How single-nucleotide polymorphisms associated with educational attainment relate to life-course development*. *Psychological science*, 27(7), 957-972.
- Bliss, C. (2018). *Social by nature : the promise and peril of sociogenomics*. Stanford University Press.
- Briley, D. A., Tropf, F. C., & Mills, M. C. (2017). *What Explains the Heritability of Completed Fertility? Evidence from Two Large Twin Studies*. *Behavior Genetics*, 47(1), 36-51.
- Bronfenbrenner, U., & Ceci, S. J. (1994). *Nature-nuture reconceptualized in developmental perspective: A bioecological model*. *Psychological review*, 101(4), 568.
- Caspi, A., McClay, J., Moffitt, T. E., Mill, J., Martin, J., Craig, I. W., ... & Poulton, R. (2002). *Role of genotype in the cycle of violence in maltreated children*. *Science*, 297(5582), 851-854.
- Cavalli-Sforza, L. L., & Bodmer, W. F. (1999). *The genetics of human populations*. Courier Corporation.
- Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). *Intelligence and educational achievement*. *Intelligence* (Norwood), 35(1), 13-21.
- DiPrete, T. A., Burik, C. A., & Koellinger, P. D. (2018). *Genetic instrumental variable regression: Explaining socioeconomic and health outcomes in nonexperimental data*. *Proceedings of the National Academy of Sciences*, 115(22), E4970-E4979.
- Domański, H. (2005). *BETWEEN STATE SOCIALISM AND MARKETS*. *European Societies*, 7(2), 197-218.
- Domingue, B. W., Belsky, D. W., Harrati, A., Conley, D., Weir, D. R., & Boardman, J. D. (2017). *Mortality selection in a genetic sample and implications for association studies*. *International journal of epidemiology*, 46(4), 1285-1294.
- Dudbridge, F. (2013). *Power and predictive accuracy of polygenic risk scores*. *PLoS genetics*, 9(3), e1003348.
- Duncan, L. E., Pollastri, A. R., & Smoller, J. W. (2014). *Mind the gap: Why many geneticists and psychological scientists have discrepant views about gene-environment interaction (G×E) research*. *American Psychologist*, 69(3), 249.
- Duster, T. (2006). *Backdoor to eugenics*. Routledge.
- Fieder, M., & Huber, S. (2007). *The effects of sex and childlessness on the association between status and reproductive output in modern society*. *Evolution and Human Behavior*, 28(6), 392-398.

- Finucane, H. K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P. R., ... & Price, A. L. (2015). *Partitioning heritability by functional annotation using genome-wide association summary statistics*. *Nature genetics*, 47(11), 1228-1235.
- Fisher, R. A. (1958). *The genetical theory of natural selection*. Рипол Классик.
- Freese, J. (2018). *The arrival of social science genomics*. *Contemporary sociology*, 47(5), 524-536.
- Freese, J., & Shostak, S. (2009). *Genetics and social inquiry*. *Annual Review of Sociology*, 35, 107-128.
- Harden, K. P., Domingue, B. W., Belsky, D. W., Boardman, J. D., Crosnoe, R., Malanchini, M., ... & Harris, K. M. (2020). *Genetic associations with mathematics tracking and persistence in secondary school*. *NPJ science of learning*, 5(1), 1-8.
- Harden, K. P., & Koellinger, P. D. (2020). *Using genetics for social science*. *Nature Human Behaviour*, 4(6), 567-576.
- Hazan, M., & Zoabi, H. (2015). *Do highly educated women choose smaller families?*. *The Economic Journal*, 125(587), 1191-1226.
- Health and Retirement Survey. (2018). *HRS Documentation Report*. https://www.rand.org/content/dam/rand/www/external/labor/aging/dataproducts/randhrs1992_2018v1.pdf
- Herrnstein, R. J., & Murray, C. (1994). *The bell curve : intelligence and class structure in American life*. The Free Press.
- Hewitt, J. K. (2012). *Editorial policy on candidate gene association and candidate gene-by-environment interaction studies of complex traits*. *Behavior genetics*, 42(1), 1.
- Hill, W. D., Davies, N. M., Ritchie, S. J., Skene, N. G., Bryois, J., Bell, S., ... & Deary, I. J. (2019). *Genome-wide analysis identifies molecular systems and 149 genetic loci associated with income*. *Nature communications*, 10(1), 1-16.
- Hoeller, Peter, Joumard, Isabelle, & Koske, Isabell. (2013). *Income inequality in OECD countries*. World Scientific Publishing Co. Pte. Ltd.
- Hopcroft, R. L. (2006). *Sex, status, and reproductive success in the contemporary United States*. *Evolution and Human Behavior*, 27(2), 104–120.
- Huber, S., Bookstein, F. L., & Fieder, M. (2010). *Socioeconomic status, education, and reproduction in modern women: an evolutionary perspective*. *American Journal of Human Biology*, 22(5), 578-587.
- Jencks, C. (1980). *Heredity, environment, and public policy reconsidered*. *American Sociological Review*, 723-736.
- Jensen, A. R. (1969). *How much can we boost IQ and scholastic achievement?*. *Harvard Educational Review*, 39(1)
- Khrantsova, E. A., Davis, L. K. & Stranger, B. E. (2019). *The role of sex in the genomics of human complex traits*. *Nature Reviews. Genetics*, 20(3), 173–190.
- Kong, A., Frigge, M., Thorleifsson, G., Stefansson, H., Young, A. I., Zink, F., Jonsdottir, G. A., Okbay, A., Sulem, P., Masson, G., Gudbjartsson, D., Helgason, A., Bjornsdottir, G., Thorsteinsdottir, U., & Zwart, J.-A.. (2017). *Selection against variants in the genome associated with educational attainment*. *Proceedings of the National Academy of Sciences - PNAS*, 114(5), E727–E732.
- Lande, R. & Arnold, S. J. (1983). *The Measurement of Selection on Correlated Characters*. *Evolution*, 37(6), 1210–1226.

- Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., ... & Cesarini, D. (2018). *Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals*. *Nature genetics*, 50(8), 1112-1121.
- Linnér, R. K., Biroli, P., Kong, E., Meddens, S. F. W., Wedow, R., Fontana, M. A., ... & Beauchamp, J. P. (2019). *Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences*. *Nature genetics*, 51(2), 245-257.
- Martin, A. R., Gignoux, C. R., Walters, R. K., Wojcik, G. L., Neale, B. M., Gravel, S., ... & Kenny, E. E. (2017). *Human demographic history impacts genetic risk prediction across diverse populations*. *The American Journal of Human Genetics*, 100(4), 635-649.
- Martin, T. C. (1995). *Women's education and fertility: results from 26 Demographic and Health Surveys*. *Studies in family planning*, 187-202.
- Mathieson, I., Day, F. R., Barban, N., Tropf, F. C., Brazel, D. M., Vaez, A., ... & Human Reproductive Behaviour Consortium. (2020). *Genome-wide analysis identifies genetic effects on reproductive success and ongoing natural selection at the FADS locus*. *bioRxiv*.
- Mills, M. C., & Rahal, C. (2019). *A scientometric review of genome-wide association studies*. *Communications biology*, 2(1), 1-11.
- Mills, M. C., & Rahal, C. (2020). *The GWAS Diversity Monitor tracks diversity by disease in real time*. *Nature genetics*, 52(3), 242-243.
- Mills, M. C., & Tropf, F. C. (2020). *Sociology, Genetics, and the Coming of Age of Sociogenomics*. *Annual Review of Sociology*, 46, 553-581.
- Mills, M. C., Tropf, F. C., Brazel, D. M., van Zuydam, N., Vaez, A., Pers, T. H., ... & Human Reproductive Behaviour Consortium. (2020). *Identification of 370 loci for age at onset of sexual and reproductive behaviour, highlighting common aetiology with reproductive biology, externalizing behaviour and longevity*. *BioRxiv*.
- Morris, T. T., Davies, N. M., & Davey Smith, G. (2019). *Can education be personalised using pupils' genetic data?* *bioRxiv*, 645218.
- Morrissey, M. B., Kruuk, L. E. B., & Wilson, A. J. (2010). *The danger of applying the breeder's equation in observational studies of natural populations*. *Journal of Evolutionary Biology*, 23(11), 2277-2288.
- Mostafavi, H., Harpak, A., Agarwal, I., Conley, D., Pritchard, J. K., & Przeworski, M. (2020). *Variable prediction accuracy of polygenic scores within an ancestry group*. *Elife*, 9, e48376.
- Myrskylä, M., Kohler, H. P., & Billari, F. C. (2009). *Advances in development reverse fertility declines*. *Nature*, 460(7256), 741-743.
- Nelson, A. (2016). *The social life of DNA: Race, reparations, and reconciliation after the genome*. Beacon Press.
- Plomin, R. (2019). *Blueprint: How DNA makes us who we are*. Mit Press.
- Plomin, R., & Von Stumm, S. (2018). *The new genetics of intelligence*. *Nature Reviews Genetics*, 19(3), 148-159.
- Popejoy, A. B., & Fullerton, S. M. (2016). *Genomics is failing on diversity*. *Nature News*, 538(7624), 161.
- 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*. *Nature genetics*, 38(8), 904-909.
- Reardon, J. (2017). *The postgenomic condition : ethics, justice, and knowledge after the genome*. The University of Chicago Press.

- Ridge, P. G., Hoyt, K. B., Boehme, K., Mukherjee, S., Crane, P. K., Haines, J. L., ... & Reitz, C. (2016). *Assessment of the genetic variance of late-onset Alzheimer's disease*. *Neurobiology of aging*, 41, 200-e13.
- Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., ... & Preisig, M. (2013). *GWAS of 126,559 individuals identifies genetic variants associated with educational attainment*. *science*, 340(6139), 1467-1471.
- Rimfeld, K., Krapohl, E., Trzaskowski, M., Coleman, J. R., Selzam, S., Dale, P. S., ... & Plomin, R. (2018). *Genetic influence on social outcomes during and after the Soviet era in Estonia*. *Nature human behaviour*, 2(4), 269-275.
- Rodgers, J. L., Kohler, H. P., McGue, M., Behrman, J. R., Petersen, I., Bingley, P., & Christensen, K. (2008). *Education and cognitive ability as direct, mediating, or spurious influences on female age at first birth: Behavior genetic models fit to Danish twin data*. *American journal of sociology*, 114(S1), S202-S232.
- Servais, M. A. (2010). *Overview of HRS public data files for cross-sectional and longitudinal analysis*. Ann Arbor, MI: Survey Research Center, Institute for Social Research.
- Sewell, W. H., & Shah, V. P. (1967). *Socioeconomic status, intelligence, and the attainment of higher education*. *Sociology of education*, 1-23.
- Skirbekk, V. (2008). *Fertility trends by social status*. *Demographic Research*, 18, 145–180.
- Tropf, F. C., Lee, S. H., Verweij, R. M., Stulp, G., Van Der Most, P. J., De Vlaming, R., ... & Mills, M. C. (2017). *Hidden heritability due to heterogeneity across seven populations*. *Nature human behaviour*, 1(10), 757-765.
- Tropf, F. C., & Mandemakers, J. J. (2017). *Is the Association Between Education and Fertility Postponement Causal? The Role of Family Background Factors*. *Demography*, 54(1), 71–91.
- Tropf, F. C., Stulp, G., Barban, N., Visscher, P. M., Yang, J., Snieder, H. & Mills, M. C. (2015). *Human Fertility, Molecular Genetics, and Natural Selection in Modern Societies*. *PloS One*, 10(6), e0126821.
- Turkheimer, E., Haley, A., Waldron, M., d'Onofrio, B., & Gottesman, I. I. (2003). *Socioeconomic status modifies heritability of IQ in young children*. *Psychological science*, 14(6), 623-628.
- Weeden, J., Abrams, M. J., Green, M. C., & Sabini, J. (2006). *Do high-status people really have fewer children?*. *Human Nature*, 17(4), 377-392.
- Zietsch, B. P., Kuja-Halkola, R., Walum, H. & Verweij, K. J. H. (2014). *Perfect genetic correlation between number of offspring and grandoffspring in an industrialized human population*. *Proceedings of the National Academy of Sciences - PNAS*, 111(3), 1032–1036.

Anhang

Tabelle 4: Multiple lineare Regression der $\sqrt{\text{Kinderzahl}}$ regressiert auf PGS Educational attainment, PGS Number of children ever born, Jahre an Bildung und Gesamteinkommen, getrennt nach Geschlecht. Standardisierte Koeffizienten und Signifikanzniveau der Ergebnisse sind angegeben.

Variable	Frauen		Männer	
	Standardisierte Koeffizienten (Beta-Wert)	Sig. p	Standardisierte Koeffizienten (Beta-Wert)	Sig. p
PGS Number of children ever born (NEB)	0,394	< 0,0001	0,492	< 0,0001
PGS Educational attainment (EA)	0,034	0,064	0,010	0,639
Zahl der Ausbildungsjahre	-0,159	< 0,0001	-0,052	0,011
Gesamteinkommen	-0,067	0,0002	0,051	0,009

Anmerkung: Die überprüfte Kontrollvariable Alter sowie die zehn "ancestry-specific principal components" (PCs) zur Vermeidung von unbeobachtetem Confounding durch Populationsunterschiede werden nicht in den Ergebnissen angeführt.

Tabelle 5: Multiple lineare Regression der $\sqrt{\text{Kinderzahl}}$ regressiert auf PGS Educational Attainment, PGS Number of children ever born, Jahre an Bildung und Rentenversicherungsvermögen, getrennt nach Geschlecht. Standardisierte Koeffizienten und Signifikanzniveau der Ergebnisse sind angegeben.

Variable	Frauen		Männer	
	Standardisierte Koeffizienten (Beta-Wert)	Sig. p	Standardisierte Koeffizienten (Beta-Wert)	Sig. p
PGS Number of children ever born (NEB)	0,373	< 0,0001	0,485	< 0,0001
PGS Educational attainment (EA)	0,044	0,023	0,007	0,661
Zahl der Ausbildungsjahre	-0,141	< 0,0001	-0,048	0,021
Rentenversicherungsvermögen	-0,170	< 0,0001	0,050	0,016

Anmerkung: Die überprüfte Kontrollvariable Alter sowie die zehn "ancestry-specific principal components" (PCs) zur Vermeidung von unbeobachtetem Confounding durch Populationsunterschiede werden nicht in den Ergebnissen angeführt.

Tabelle 6: Lineare Regression der $\sqrt{\text{Kinderzahl}}$ regressiert auf PGS Number of children ever born getrennt nach Geschlecht. Standardisierte Koeffizienten und Signifikanzniveau der Ergebnisse sind angegeben.

Variable	Frauen		Männer	
	Standardisierte Koeffizienten (Beta-Wert)	Sig.	Standardisierte Koeffizienten (Beta-Wert)	Sig. p
PGS Number of children ever born (NEB)	0,356	< 0,0001	0,371	< 0,0001

Anmerkung: Die überprüfte Kontrollvariable Alter sowie die zehn "ancestry-specific principal components" (PCs) zur Vermeidung von unbeobachtetem Confounding durch Populationsunterschiede werden nicht in den Ergebnissen angeführt.

Tabelle 7: Lineare Regression des Gesamteinkommens regressiert auf den PGS Educational attainment. Standardisierte Koeffizienten und Signifikanzniveau der Ergebnisse sind angegeben.

Variable	Standardisierte Koeffizienten (Beta-Wert)	Sig. p
PGS Educational Attainment (EA)	0,105	< 0,0001

Abhängige Variable: Gesamteinkommen

Anmerkung: Die überprüfte Kontrollvariable Alter sowie die zehn "ancestry-specific principal components" (PCs) zur Vermeidung von unbeobachtetem Confounding durch Populationsunterschiede werden nicht in den Ergebnissen angeführt.