

## Article

# Exploring the Lifetime Effect of Children on Wellbeing Using Two-Sample Mendelian Randomisation

Benjamin Woolf <sup>1,2,3,\*</sup> , Hannah M. Sallis <sup>2,4,†</sup> and Marcus R. Munafò <sup>1,2,†</sup><sup>1</sup> School of Psychological Science, University of Bristol, Bristol BS8 1TU, UK<sup>2</sup> MRC Integrative Epidemiology Unit, University of Bristol, Bristol BS8 2BN, UK<sup>3</sup> MRC Biostatistics Unit, University of Cambridge, Cambridge CB2 1TN, UK<sup>4</sup> Centre for Academic Mental Health, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol BS8 2BN, UK

\* Correspondence: benjamin.woolf@bristol.ac.uk

† These authors contributed equally to this work.

**Abstract:** Background: Observational research implies a negative effect of having children on wellbeing. Objectives: To provide Mendelian randomisation evidence of the effect of having children on parental wellbeing. Design: Two-sample Mendelian randomisation. Setting: Non-clinical European ancestry participants. Participants: We used the UK Biobank (460,654 male and female European ancestry participants) as a source of genotype-exposure associations, the Social Science Genetics Consortia (SSGAC) (298,420 male and female European ancestry participants), and the Within-Family Consortia (effective sample of 22,656 male and female European ancestry participants) as sources of genotype-outcome associations. Interventions: The lifetime effect of an increase in the genetic liability to having children. Primary and secondary outcome measures: The primary analysis was an inverse variance weighed analysis of subjective wellbeing measured in the 2016 SSGAC Genome Wide Association Study (GWAS). Secondary outcomes included pleiotropy robust estimators applied in the SSGAC and an analysis using the Within-Family consortia GWAS. Results: We did not find strong evidence of a negative (standard deviation) change in wellbeing ( $\beta = 0.153$  [95% CI:  $-0.210$  to  $0.516$ ] per child parented). Secondary outcomes were generally slightly deflated (e.g.,  $-0.049$  [95% CI:  $-0.533$  to  $0.435$ ] for the Within-Family Consortia and  $0.090$  [95% CI:  $-0.167$  to  $0.347$ ] for weighted median), implying the presence of some residual confounding and pleiotropy. Conclusions: Contrary to the existing literature, our results are not compatible with a measurable negative effect of number of children on the average wellbeing of a parent over their life course. However, we were unable to explore non-linearities, interactions, or time-varying effects.

**Keywords:** Mendelian randomisation; wellbeing; family size

**Citation:** Woolf, B.; Sallis, H.M.; Munafò, M.R. Exploring the Lifetime Effect of Children on Wellbeing Using Two-Sample Mendelian Randomisation. *Genes* **2023**, *14*, 716. <https://doi.org/10.3390/genes14030716>

Academic Editors: Chenglong Yu, Shuo Chen, Shuai Li and Hsin-Hsiung Huang

Received: 19 January 2023

Revised: 6 March 2023

Accepted: 7 March 2023

Published: 14 March 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

A well-replicated [1–8] but contested [9–14] finding in the observational quantitative social science literature is a negative association between having children and subjective wellbeing in English-speaking countries. Although many studies have failed to replicate the finding in non-English-speaking countries [15–17], others have [18,19]. For example, Novoa and colleagues found a negative association between having children and subjective wellbeing in Chile [20]. Matters are further complicated by non-linearities depending on the age at which wellbeing is assessed in the parents. According to some of the studies, having children is not negatively associated with wellbeing when measured in geriatric populations, especially for those with a lower socioeconomic position (SEP), despite a negative association when measured in younger age groups [18,20–23]. This is possibly because of the increased social support that children can provide to their parents in old age [2,14,21,22].

Because these findings come from mostly cross-sectional studies, they should be treated with caution. For example, most of the studies cited above adjusted for only a few potential confounding variables (e.g., age, sex, and socio-economic position), if any, raising a question of residual confounding. Indeed, Deaton and Stone found that the choice of covariates adjusted for could produce radically different conclusions [15]. There will also likely be underlying psychological differences between adults who choose to have children and those who choose not to have children. Since the examined studies did not adjust for these variables, if these psychological differences also directly influence wellbeing, then they would be an additional source of unmeasured confounding. Relatedly, because most of the literature is derived from correlational surveys, these studies cannot ascertain the direction of effect [6]. Finally, cross-sectional studies cannot determine if the apparent change in effect with age is a cohort effect instead, although the corroboration of this finding by some prospective studies makes this less plausible [8]. As observational studies, these cohort studies are likely to still suffer from residual confounding.

Because of the higher risk of bias in traditional observational studies, it is becoming increasingly common in social epidemiology and econometrics to triangulate evidence from conventional observational studies with quasi-experimental designs [24]. One such design is Mendelian randomisation (MR) [25–27]. In a randomised controlled trial, participants are randomised to an intervention or control arm and followed up for a certain period. Because genetic variants are inherited at random, comparing the outcome status of an individual with and without a causal variant for the exposure is essentially analogous to a clinical trial [28]. In addition, because our genotype is fixed at conception, MR estimates are robust to reverse causation. This also means that any effect estimate derived from MR studies should be interpreted as the lifetime effect of the exposure. Applications of MR have traditionally focused on biomedical exposures where the analogy between MR and randomised controlled trials for a pharmacological intervention is strong because most drugs target proteins, which are the proximal product of genes [29].

MR has been gaining popularity as a method for answering causal questions in medicine, psychology and the social sciences in recent years [30,31]. The ‘social environment’, defined as the behaviour (and consequences) of those around us, is a potentially important cause of ill health and is a psychological mechanism. For example, one person’s smoking could cause those around them to smoke as well (‘social smoking’). Psychiatric genetics and evolutionary theory both suggest that the behaviours of those close by will associate with that person’s genotype. These are postulated to occur through gene–environment correlation, which may be active (when someone’s heritable phenotype influences someone to select into an environment) or evocative (when heritable phenotype evokes a change in the environment) and through the extended phenotype (when genes associate with the environment due to the effects of the heritable phenotype) [32,33]. For example, someone’s genetic predisposition to smoke may associate with their exposure to passive/environmental tobacco smoke, either by provoking social smoking or by homophily (the tendency of people to be friends with similar people). However, to the best of our knowledge, MR has not been used to explore the effects of social environmental exposures. This is, in part, because it is conceptually less clear as to how the potential effect of a genetic variant, which robustly increases the probability of a social environmental exposure—such as a traffic accident—by a very small amount, is genuinely equivalent to an exposure such as being hit by a lorry. Children, on the other hand, are a part of one’s social environment, but can plausibly be studied with a genetic design such as MR because having children is the primary endpoint through which evolution by natural selection occurs, and hence should be influenced by genetics [34,35]. However, the effects that children have on their parents after their birth is a social, rather than a direct biological, effect.

Therefore, exploring the effect that children have on their parent’s wellbeing not only answers a question of societal importance, but can also support the use of leveraging an active gene–environmental correlation within an MR study design to study the causal effect

of environmental exposures. We therefore used a two-sample MR to explore the lifetime effect of having children on wellbeing.

#### **Strengths and limitations of this study**

- Mendelian randomisation (MR) is a natural experiment that is theoretically robust for confounding and reverse causation.
- We were able to use two negative control analyses to explore the robustness of our study to two potential sources of residual confounding (populations structure and passive gene–environment correlation).
- We additionally used pleiotropy robust estimates (such as MR-PRESSO, MR-Egger, weighted median, and weighed mode) to explore if our result was affected by the direct effects of the genetic variants on the outcome not mediated by the exposure.
- Because we used summary data, we were unable to explore interactions or non-linear, time-varying, or time sensitive effects.

## **2. Methods**

### *2.1. Study Design*

We performed a two-sample MR analysis to explore the lifetime effect that children have on parental wellbeing. Specifically, we used the UK Biobank (UKB) as a source of genetic instruments [36,37], as well as their weights, for the number of children an individual has, and we used the 2016 Social Science Genetics Consortium (SSGAC) Genome Wide Association Study (GWAS) meta-analysis of subjective wellbeing as a source of instrument–outcome associations [38].

### *2.2. Data Sources*

#### *2.2.1. UK Biobank (UKB)*

The UKB is a large (~500,000 participants) population cohort study in the UK. Members of the public between the ages of 38 and 73 and who lived within 22 miles of an assessment centre were invited to participate from 2006 to 2010. Approximately 9.2 million individuals were invited to take part, with around 6% participating in the baseline assessment. The sample was 55% female, and predominantly of European ancestry (96%). The study design, participants and quality control (QC) methods have been described in full elsewhere [36]. UKB received ethics approval from the North West Multi-Centre Research Ethics Committee (REC reference 11/NW/0382). All participants provided written informed consent to participate in the study.

#### *2.2.2. Social Science Genetics Consortia (SSGAC)*

Data on subjective wellbeing was extracted from the 2016 SSGAC subjective wellbeing GWAS (OpenGWAS ID: ieu-a-1009) [38]. This was a meta-analysis of 298,420 individuals from 59 studies. Samples included men and women, mostly of European descent, living in Europe, North America, or Australia.

#### *2.2.3. Within Family Consortium (WFC)*

Data on wellbeing was also taken from the 2022 Within Family Consortium (WFC) GWAS (OpenGWAS ID: ieu-b-4851) [39]. This GWAS used the genetic overlap of relatives to adjust for which SNPs were inherited and therefore fully removed most plausible sources of confounding, including ancestry and genetic nurture [40]. The consortium combines data on almost 160,000 pairs of relatives from 17 cohorts. The GWAS itself had an effective sample size of 22,656 (male and female) individuals of European ancestry. A more detailed description of the individual cohorts included can be found in the paper's Supplementary Materials [39]. To avoid confusion, the WFC does not estimate between-relative effects, but instead the association between one person's genotype and another person's phenotype. The WFC uses genetic data from biological relatives to then remove biases in population genetics studies such as assortative mating and population structure.

### 2.3. Phenotyping

#### 2.3.1. UKB

Information on the number of (biological) children (OpenGWAS ID: ieu-b-4760) a participant had, number of full sisters (UKB ID: 1883, OpenGWAS ID: ukb-b-5593), number of full brothers (UKB ID: 1873, OpenGWAS ID: ukb-b-4263), general happiness (UKB ID: 20458, OpenGWAS ID: ukb-b-4062), number of older siblings (UKB ID: 5057, OpenGWAS ID: ukb-b-1997), and hair colour (UKB ID: 1747, Open GWAS IDs: ukb-d-1747\_5, ukb-d-1747\_4, ukb-d-1747\_3, ukb-d-1747\_1, ukb-d-1747\_2, ukb-d-1747\_6) were collected through a questionnaire asked either during the initial visit to an assessment centre or, in the case of general happiness, in an online follow up. Number of biological children was measured by asking men (UKB ID: 2405) “How many children have you fathered?”, and women (UKB ID: 2734) “How many children have you given birth to? (Please include live births only)”. The exact questions asked for the other phenotypes are provided in the Supplementary Materials.

#### 2.3.2. SSGAC

This study included measures of life satisfaction, positive affect, or both in the GWAS. The specific questionnaires used for phenotype subjective wellbeing in each sample are described in the Supplementary Materials [38]. These were standardised for the meta-analysis. In most of the samples used in the SSGAC GWAS, a 1 to 2 standard deviation change is equivalent to a 1 unit increase on a 5-level psychometric question. For example, the 23andMe study asked participants to rate from very dissatisfied with their life (score = 0) to very satisfied with it (score = 5) and had a standard deviation of 1. This means that a one standard deviation increase would be the same as going from very dissatisfied to somewhat dissatisfied. Supplementary Tables S1–S3 provide the gene-exposure and gene-outcome associations used in this study.

#### 2.3.3. WFC

All participating cohorts measured wellbeing using a questionnaire. Wellbeing measures were standardised prior to the meta-analysis. More details on phenotyping are provided in the Supplementary Materials of the original paper [39].

### 2.4. Statistical Analysis

#### 2.4.1. Overview of the Analysis

The primary analysis was an inverse variance weighted meta-analysis of the Wald ratio using independent genome-wide significant ( $p < 5 \times 10^{-8}$ ) SNPs for number of children identified in UKB as instruments [37]. The Wald ratios were defined as the variant's association with wellbeing divided by the variant's association with number of children.

We additionally used a number of sensitivity analyses, including (a) five pleiotropy robust estimators (MR-Egger, MR-RAPS, MR-PRESSO, weighted median, and weighted mode), (b) two sets of negative controls (hair colour, and number of parental siblings) as a falsification test for the presence of residual confounders of the instrument–outcome association, (c) the Within-Family Consortium (WFC) as a more robust (but less well powered) outcome GWAS, and (d) a less stringent  $p$ -value threshold ( $p < 5 \times 10^{-6}$ ) for selecting SNPs to increase power. More details can be found in the Supplementary Materials.

#### 2.4.2. Instrument Construction

Genetic instruments were selected using a statistical criterion of having a genome-wide significant association with the exposure ( $p < 5 \times 10^{-8}$ ). We additionally clumped the variants using an  $r^2$  of 0.001 and KB of 10,000, thereby ensuring that the instruments were independent of each other. Genetic variants that have their gene–exposure association estimated in the same dataset used to select instruments can suffer from a bias called ‘Winner’s Curse’. This occurs because variants can meet the statistical criteria due to having a genuine association or because they, by chance, have an unusually large amount of noise. This results in variants appearing to have a larger association than they actually do. We

therefore further filtered these variants using the False discovery rate Inverse Quantile Transformation (FIQT) winners curse correction developed by the SSGAC [41]. This uses an analogy between multiple testing and winners curse to apply an easy-to-implement correction to effect estimates.

We used the TwoSampleMR R package to harmonise the two GWASs. Palindromic SNPs were only excluded if their allele frequency could not be used to infer which strand was positive. In cases where SNPs were missing in the outcome dataset, we used TwoSampleMR to automatically impute LD proxy variants, using an  $r^2$  of 0.8 from the European subsample of the 1000 genomes project.

MR assumes that the causal pathways go from the genetic variant to the exposure to the outcome. However, when selecting genetic instruments using a data-driven method, as we did, it is possible to select some SNPs wherein the causal path is from the variant to the outcome to the exposure. When the exposure and outcome GWASs are of similar power, one should expect SNPs to explain a larger proportion of the variance in the more proximal GWAS than the more distal one. Steiger filtering, which we applied, uses this logic as a method of removing SNPs that are more proximal to the outcome than the exposure.

#### 2.4.3. Statistical Methods

The primary MR estimator in this study was the Wald ratio. This is defined by the variant–outcome association divided by the variant–exposure association. Because both the variant–exposure and variant–outcome associations were derived from linear models, MR analysis also assumed a linear model.

We then used six methods of meta-analysing for the MR estimates of each SNP: IVW, MR-Egger, weighted median, weighted mode, MR-RAPS, and MR-PRESSO. The IVW estimate will return the true effect if all the IV assumptions are valid. The other five ‘pleiotropy robust’ methods can return the true effect if some of the instruments are invalid, but they have reduced power. Specifically, the weighted mode assumes that the modal effect size is a valid estimate of the true effect size, while weighted median assumes that at least half of the SNPs are valid. IVW can be thought of as a regression of the variant–outcome association on the variant–exposure association, with the intercept fixed at zero [42]. MR-Egger extends this model to allow for a non-zero intercept. If we assume that the variant–exposure effect size is independent of the size of any bias (such as a pleiotropic effect), then the biasing pathways should impact the intercept but not the slope. This assumption is called the INSIDE assumption. Additionally, MR-Egger assumes that there is no measurement error in the exposure GWAS (called the NOME assumption) [43,44]. MR-RAPS assumes that pleiotropic SNPs are outliers (modelled by a random effects parameter with a mean of zero) and therefore down weights them in a random effect meta-analysis. We implemented MR-RAPS using a square error loss function, accounting for overdispersion (i.e., systematic pleiotropy) [45]. MR-RAPS is robust to both balanced pleiotropy in non-outlier SNPs and weak instrument bias. MR-PRESSO first runs a global test for pleiotropy, and then tests for outliers. It then excludes outliers before running a meta-analysis and testing for a difference between the MR estimates with and without the outliers. MR-PRESSO assumes INSIDE and that >50% of SNPs are valid instruments [46].

#### 2.4.4. Assumptions of the Analysis

Two-sample MR is an extension of MR to a summary data setting. MR is itself an extension of Instrumental Variables (IV) analysis for genetics. IV makes three assumptions: (1) relevance—that the variant is robustly associated with the exposure, (2) independence—that there are no variant–outcome confounders, (3) exclusion restriction—that the variant causes the outcome only through the exposure. For the point estimate to be interpretable, IV analysis additionally has to make an identification assumption. Here, we make the NO Simultaneous Heterogeneity (NOSH) assumption [47]. This assumption is valid if the causes of heterogeneity in the variant–exposure association are not the same as the causes

in the variant–outcome associations, or if either of these associations are homogeneous. Finally, two-sample MR makes two additional assumptions: (1) That the GWASs come from homogeneous populations. This is required for the MR estimate to be meaningful. (2) That there is no sample overlap. The effect of the second assumption is to make weak instrument bias deflationary [48–50].

Because two-sample MR uses summary data from previously conducted GWASs, if these studies assume, as they typically do, a linear effect, then it is not possible to explore non-linearity using summary data. Traditional IV estimators, such as the Wald ratio used in the two-sample MR, can still provide a valid estimate of the average causal effect in the presence of non-linearities [51].

#### 2.4.5. Assessment of Assumptions

Weak instrument bias is inversely proportional to the F-statistic of the variant–exposure association, with a common threshold of ten being required. We therefore calculated the mean F-statistic for the variant–exposure association, as well as the F-statistic for each SNP. MR assumes that there are no confounders of the variant–outcome association.

Although the other assumptions made in an MR analysis are not provable, it is possible to run falsification tests on many of them. To ensure that population structure was controlled for, we checked that the GWASs either used BOLT-LMM, which adjusts the GWAS for the entire genetic relationship matrix [52], or adjusted for the principal components of the genetic relationship matrix. LD score regression (LDSC) can be used to explore the residual population structure in GWASs that do not use linear mixed models such as BOLT-LMM. Specifically, if the LDSC intercept is very different from one, this implies the presence of a residual population structure [53]. We also ran a set of sensitivity analyses, described below, as falsification tests for residual gene–outcome confounders.

Horizontal pleiotropy occurs when a genetic variant associates with two phenotypes for independent reasons. This can cause a violation of the exclusion restriction assumption. However, if horizontal pleiotropy is present, the exact pathway should be different for each variant. It has therefore been argued that, if it occurs, it should create heterogeneity in the MR estimates, and its presence can therefore be tested using a heterogeneity statistic. The presence of horizontal pleiotropy was also visually explored using a funnel plot.

The NOME assumption can be tested by checking that the  $I^2$  statistic for the gene–exposure association is greater than 90% [43]. Since this is a measure of the variability in the variant–exposure association, we will refer to it as the  $I^2_{GX}$  statistic for clarity.

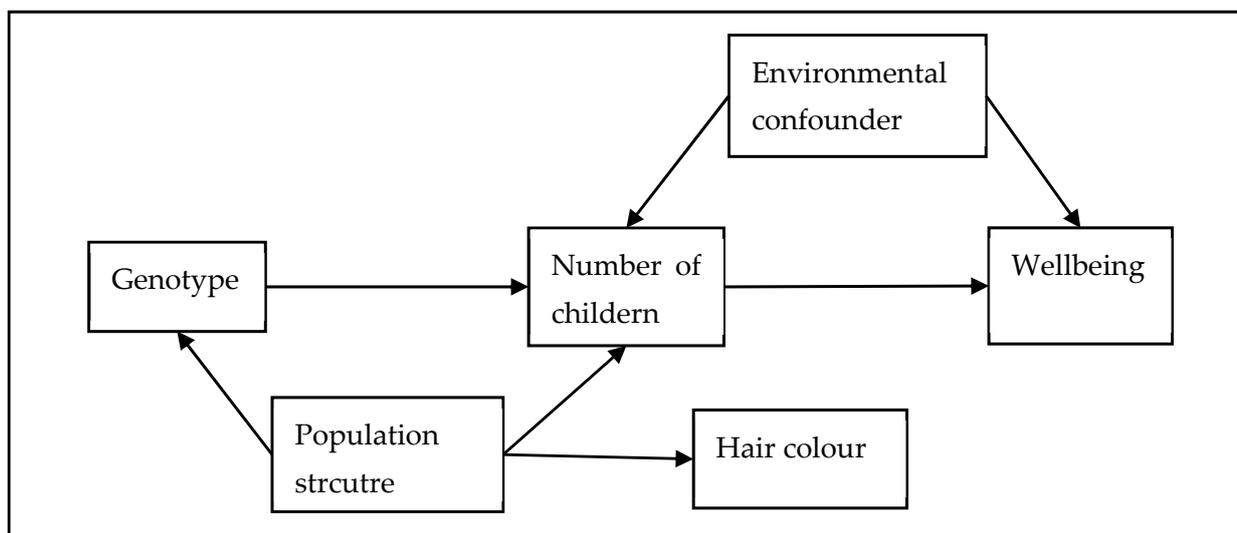
We checked sample overlap by checking which samples were reported as being included in each of the GWAS consortia in the respective publications. Two-sample MR studies generally validate the assumption that samples are drawn from the same population by checking that the studies are demographically similar [50]. However, in situations where their instrument–exposure or instrument–outcome association has been measured in both the exposure and outcome samples, it can be possible to validate this assumption quantitatively. One would expect to find only chance differences in estimates drawn from the same population [54,55]. Therefore, on top of comparing demographic information and because information on happiness from the UKB was used in the SSGAC GWAS, we checked that the average-difference SNP estimates from the UKB compared to the SSGAC and WFC were approximately zero. Because the differences in the precision of the measure of wellbeing between the UKB and SSGAC could introduce heterogeneity into the estimation of the difference in SNP effects, we chose a random-effects meta-analysis as the primary estimator.

### 2.5. Sensitivity and Additional Analyses

#### 2.5.1. Negative Controls

We used two sets of negative control outcomes to explore two potential sources of confounding. Firstly, hair colour is known to vary by ancestry in the UK/European population, but *prima facie* should not have a direct causal relationship with the number of

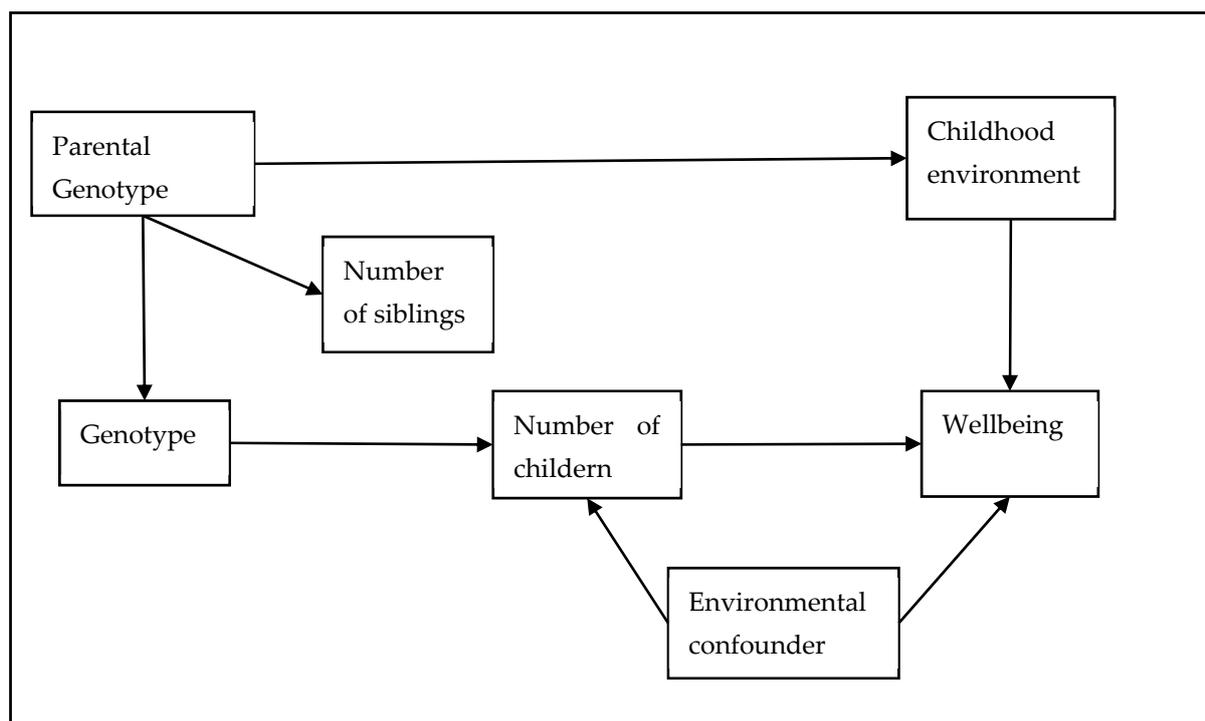
children we have [56]. It can therefore be used as a negative control outcome for population structure (see Figure 1 for a visual representation). Secondly, wellbeing may be affected by our developmental environment. However, the number of children we have is affected by our parental genotype (through inheritance). The parental genotype will also influence our developmental environment because of the effects that the parental genotype will have on the parental phenotype and therefore on the developmental environment (i.e., through genetic nurture, see Figure 2 for a visual representation). To test if genetic nurture is a residual confounder, we used the number of siblings as a negative control outcome. The number of siblings an individual has is caused by parental genotype but is unlikely to be caused by the number of children that the individual has. Therefore, the most likely explanation of an MR association with number of siblings is residual confounding due to genetic nurture. We ran the negative controls only using an IVW estimator because it is the most efficient estimator and thus the most likely to detect an association if there is one. In addition, we wanted an estimator which was *not* robust to pleiotropy because a pleiotropic association between the genetic instrument and the negative control outcomes (i.e., which is not mediated by number of children) is still evidence of an association between the instrument and a confounder of the instrument–outcome association, and hence a violation of the independence assumption.



**Figure 1.** Directed acyclic graph for the hair colour negative control outcome. Hair colour is hypothesised to be associated with population structure, but to not have a direct causal link to either the genetic instruments or wellbeing. Therefore, any association between the genetic instruments and hair colour will be due to a residual confounding effect of population structure.

### 2.5.2. WFC GWAS

Family data has been proposed as a way of eliminating the risk of confounding in MR studies [40]. Potential violations of the independence assumption, such as population structure and genetic nurture, occur because the distribution of SNPs in a population GWAS is only approximately random. By conditioning on the parental genotype, however, the back door path used by these biases is blocked. Although these GWASs are less biased, the use of non-independent observations means that they need much larger samples than a population GWAS to achieve the same level of power. We therefore used the WFC GWAS of well-being as an additional sensitivity analysis to explore the robustness to potential confounders.



**Figure 2.** Directed acyclic graph for the number of siblings negative control outcome. The parental genotype determines not only their children’s genotype, but also, via the parental phenotype, the environment in which the children grew up. If someone’s childhood environment influences their wellbeing later in life, then the parental genotype would confound any association between an individual’s genetic instruments and wellbeing. Because, in the UK, parents typically stop having new children before their children start having children, it is unlikely that a child having children will influence his or her parents to have more children. Hence, the number of siblings an individual has should not be caused by an individual’s genetic liability to having children, but will be influenced by the parental liability to having children. Therefore, any association between an individual’s genetic liability to having children and the number of siblings they have would be an indicator of residual confounding due to the parental genotype.

### 2.5.3. Less stringent SNP Selection

Power in two-sample MR studies is a function of instrument strength, the precision of the outcome GWAS, and the number of instruments. Since we are limited to using pre-collected data, the outcome GWAS’s precision cannot be varied. However, a  $p$ -value threshold of  $5 \times 10^{-6}$  equates to an F-statistic of approximately 10 and should therefore not lead to weak instruments, but because it is 100-fold larger, it should increase the number of SNPs used in the analysis. We therefore also used SNPs with an indicative association ( $p < 5 \times 10^{-6}$ ) with the exposure to explore how sensitive the primary analysis was to a potentially better-powered set of instruments.

### 2.5.4. Leave-One-Out Analysis

We additionally used a ‘leave-one-out’ analysis and the MR-PRESSO outlier test, as part of our MR-PRESSO analysis, to explore if any of the variants were outliers and had a disproportionate effect on the overall IVW estimate. The ‘leave-one-out’ sensitivity analysis works by excluding each SNP in turn and running the IVW analysis without the excluded SNP. We then visually inspected the leave-one-out plot for indication that a single SNP had a disproportionate effect on the MR results. The MR-PRESSO outlier test explores whether the observed values are different from the expected values based on a regression model of the variant–outcome associations on the variant–exposure associations. Additional information, including details on genotyping, instrument construction, and MR

estimators, can be found in the Supplementary Materials. This manuscript conforms to the STROBE-MR guidelines [57].

### 2.5.5. Bidirectional Analysis

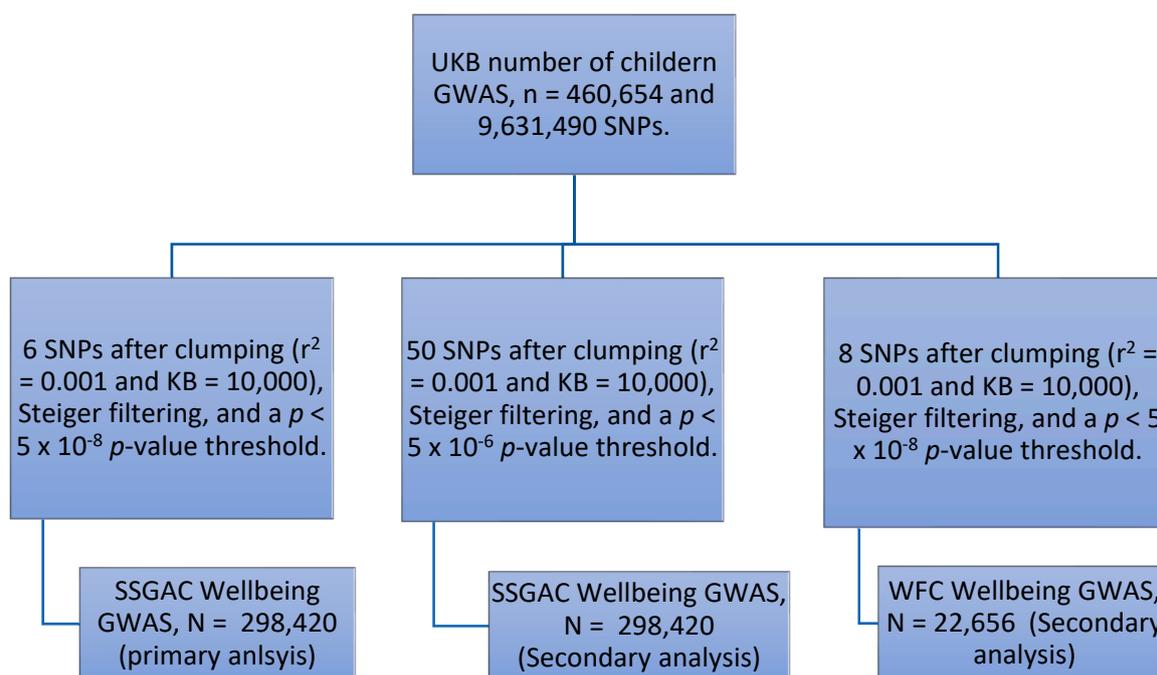
In addition to the analyses planned in our protocol, we included a bidirectional analysis in order to explore the possibility that reverse causation might explain the observational results. Specifically, we selected variants associated with well-being in the SSGAC GWAS at a  $5 \times 10^{-6}$   $p$ -value threshold, and we used the FIQT Winner's Curse correction. After harmonizing the UKB number of children GWAS with these variants, we applied Steiger filtering and estimated the MR estimates using MR-RAPS with overdispersion and a Huber loss function. This estimator was chosen because it is reasonably robust to both weak instrument bias and pleiotropy.

## 3. Results

### 3.1. Descriptive Data

#### 3.1.1. Number of Participants and SNPs in Each Stage

The UKB exposure GWAS had information from over 460,000 participants on almost 10 million SNPs. The primary outcome analysis used information on six of these SNPs, which were significant genome-wide for number of children from 298,420 participants from the SSGAC. This was increased to 50 SNPs by using a  $5 \times 10^{-6}$   $p$ -value threshold. The WFC outcome data used information on eight genome-wide significant SNPs with an effective sample size of 22,656 participants (Figure 3).



**Figure 3.** Flow chart of SNPs and participants.

#### 3.1.2. Two-Sample MR Specific Assumptions

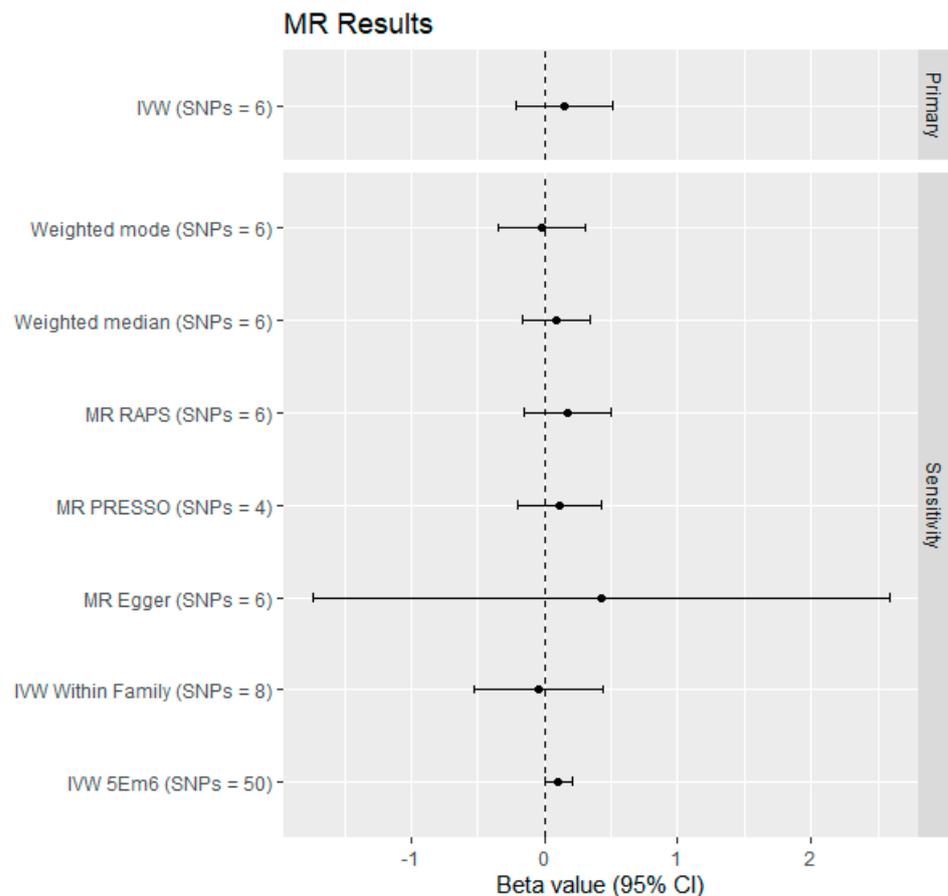
Both outcome samples had some overlap with the UKB. The SSGAC does not state how many UKB participants were included; however, around 157,000 UKB participants provided information on the measure of general happiness used by the SSGAC, which entails a maximum sample overlap of around 53% for the SSGAC and 34% for the UKB. The UKB also contributed around 4250 sibships to the WFC wellbeing GWAS, which equates to around 19% of the WFC sample and 2% of the UKB sample.

Because all GWASs were drawn from European populations of both males and females with some overlapping participants, it seems likely that the samples can all be treated as

coming from the same population. We also found no evidence of a difference in SNP effect estimates for the SNP–outcome associations in the UKB and the WFC or the SSGAC consortium using either a  $5 \times 10^{-6}$   $p$ -value or  $5 \times 10^{-8}$   $p$ -value threshold, which further supports this conclusion (Supplementary Figures S1–S3) [58].

### 3.2. Main Results

The primary IVW estimate does not indicate strong evidence of a negative (standard deviation) change in wellbeing ( $\beta = 0.153$  (95% CI: 0.210 to 0.516) per child parented (Figure 4 and Supplementary Figure S4).



**Figure 4.** Forest plot of the primary and sensitivity analyses.

### 3.3. Assessment of Assumptions

#### 3.3.1. Weak Instrument Bias and NOME

For the primary analysis, the F-statistic was 49, and the  $I^2_{GX}$  for the instrument–exposure association was 98%. These both imply that there would be around a 2% error in the MR estimates due to weak instrument bias. For the analysis using the WFC, the F-statistic was 44, and the  $I^2_{GX}$  was 98%. For the analysis using a less stringent  $p$ -value, the F-statistic was 25, and the  $I^2_{GX}$  was 96%.

#### 3.3.2. Heterogeneity and Exclusion Restriction Violations

The Cochran Q statistic for the Wald ratios of the primary analysis was 24.59 ( $p < 0.001$ ), and the  $I^2_{GX}$  for the Wald ratio was 80%. Together with the asymmetric funnel plot (Supplementary Figure S5) and the MR-PRESSO global test for outliers ( $p = 0.006$ ), this implies the presence of some pleiotropic SNPs. However, the Egger intercept was  $-0.004$  (SE = 0.017,  $p = 0.815$ ). Similar results were found for the secondary analyses (Supplementary Table S4).

### 3.4. Sensitivity and Additional Analyses

#### 3.4.1. Pleiotropy Robust Estimators

The pleiotropy robust estimates were mostly similar to the IVW estimate, although generally slightly deflated (Figure 4 and Supplementary Table S4). The exception to this was the MR-Egger estimate, although the wide 95% confidence interval (which overlaps with the IVW estimate) implies that this could be due to a lack of precision. In addition, the MR-RAPS estimate was slightly inflated, probably because RAPS is robust to both moderate amounts of weak instrument bias and pleiotropy.

#### 3.4.2. Negative Controls

The negative control outcome analysis did not find any evidence of an association of the instruments with hair colour; however, there was evidence of an association with two out of three of the sibling questions ( $p < 0.001$  for number of full brothers and  $p = 0.005$  for number of full sisters), implying the possibility of some residual confounding (Supplementary Table S5).

#### 3.4.3. WFC Outcome

Consistent with the negative control analysis, the WFC secondary analysis showed deflated point estimates compared to when using the SSGAC outcome GWAS (Supplementary Table S3). For example, the IVW estimate was  $-0.049$  (95% CI:  $-0.533$  to  $0.044$ ).

#### 3.4.4. Less Stringent SNP Selection

The standard error of the IVW estimate when using a  $5 \times 10^{-6}$  threshold was more than three times smaller than when using the more traditional  $5 \times 10^{-8}$  threshold. Supplementary Table S4 presents the results of this sensitivity analysis for all estimators. MR-RAPS in particular, as a weak instrument robust estimator, still did not find strong evidence against the null hypothesis ( $\beta = 0.082$ ,  $SE = 0.054$ ,  $p = 0.134$ ).

#### 3.4.5. Leave-One-Out Analysis and MR-PRESSO Outlier Test

The MR-PRESSO outlier test for the primary analysis identified rs10270358 and rs72687493 as outliers. An exploratory search of Phenoscanner showed no phenotypes associated with rs72687493 but found that rs10270358 is associated with seeing a doctor for anxiety or depression, as well as chronic disability/infirmity, both of which could reduce wellbeing [59]. However, these SNPs did not seem to introduce a bias in the leave-one-out analysis (Supplementary Figure S6), and the outlier test did not detect any outliers in the secondary analyses.

#### 3.4.6. Bidirectional MR

Our bidirectional analysis found a negative association between wellbeing and number of children ( $\beta = -0.069$ , 95% CI  $-0.010$  to  $-0.128$ ), implying that having lower wellbeing causes people to have more children.

## 4. Discussion

Contrary to the existing observational literature, our results do not imply that children have a detrimental effect on parental wellbeing. If we assume the measures of wellbeing in the SSGAC are on a ratio scale, and that people score the nearest category to what they feel, then our 95% confidence interval would be compatible with either no effect or as much as a one-unit increase on this scale for each child someone has (e.g., from neither satisfied nor dissatisfied to somewhat satisfied), but incompatible with a measurable negative change in subjective wellbeing.

However, our additional and sensitivity analyses imply that our primary point estimate may overestimate the true effect. Our negative control analysis found that our instrument was associated with the number of siblings, and that the point estimate was

deflated when using the WFC outcome GWAS, implying the presence of residual confounding due to genetic nurture. Likewise, the heterogeneity statistics implied the presence of residual confounding, while most pleiotropy robust estimators were again deflated. It is therefore likely that the true effect, if existent, will be smaller than a measurable change (i.e., a one-unit increase or decrease in a 5-level psychometric question) in well-being for every child a parent has.

Our bidirectional analysis indicated that having lower wellbeing causes people to have more children. This implies that reverse causation could account for some of the observational findings of a negative association between wellbeing and number of children. However, because of the small effect size (1 standard deviation decrease in wellbeing resulting in 0.07 more children), this is unlikely to entirely explain the observed association.

Our study's results appear to contradict a recent study by Giannelis and colleagues (2021) [60]. They used observational data from the UKB and two-sample MR (using the UKB and Psychiatric Genomics Consortium) to explore the effects of cohabitation and having children on depression. Their headline results appear to imply that children increase the risk of depression using IVW. Since depression plausibly has a negative effect on wellbeing, this finding is superficially contradictory to ours [61]. However, none of the directional pleiotropy robust estimators show deflated estimates, and they do not find evidence of an effect. Since they use a liberal  $p$ -value threshold to select SNPs, this is less likely to represent low power. Thus, as with our study, it is likely that IVW has overestimated the true effect.

#### 4.1. Pre-Specified Interpretation

In the study protocol, we pre-specified how we would interpret the findings of our sensitivity analyses. Specifically:

##### 4.1.1. Pleiotropy

Because the indicators for the presence of pleiotropy (such as the  $I^2_{GX}$  and Cochrane Q statistics for the Wald ratio, funnel plot and MR-RAPS) all indicated the presence of pleiotropy, and because the 'pleiotropy robust' estimators generally had deflated estimates compared to the IVW estimate, it seems likely that the IVW estimates were inflated by some residual pleiotropy.

##### 4.1.2. Residual Confounding

The association of the instruments with two out of three of the sibling negative controls, combined with the change in estimate from the WFC GWAS, implies that there was some inflation due to residual confounding from genetic nurture in the primary IVW estimate.

##### 4.1.3. Low Power

The number of SNPs increased almost nine-fold in this secondary analysis when compared to the primary one. This resulted in a three-fold decrease in the size of the standard error (0.185 to 0.053) for the IVW estimate, and implies that there is also a large amount of residual random error in the estimates. However, the point estimates in this analysis were generally deflated when compared to the primary analysis. This should be in part explained by the approximate halving of the F-statistic.

#### 4.2. Generalisability

One possible explanation for the discrepancy between the observational and MR estimates, other than residual confounding in the observational studies, is that the target estimates are not directly comparable. If we assume that age does not modify the variant-exposure association, then MR estimates should be interpreted as the average effect of the exposure on an outcome over the entire lifetime up to recruitment (typically in their late 50s at recruitment for the UK Biobank, which is the single largest study in the SSGAC) [62]. This means that the transient effects of having children on wellbeing, such as the stress of looking after a new-born baby, may not be detectable in a typical MR design. Some studies

found that having children was beneficial to parental wellbeing in old age [18,20–23]. One possible explanation for the discrepancy between our results and the existing literature's results would therefore be that the transient negative effect is counterbalanced by the later positive ones—resulting in an average effect close to zero. Likewise, there could be effects in age groups older than those included in our study. Although methods for addressing time-varying exposures are currently being developed [63], there is still no consensus on how to best estimate transient effects within an MR framework. We are therefore unable to empirically explore this interpretation further here. MR still produces a valid test for the average lifetime effect of exposure to genetically predicted levels of the exposure [64]. We therefore believe that our results are a valid, if blunt, measure of the total longitudinal effect up to recruitment. However, further quasi-experimental studies should also be leveraged to explore short term effects. For example, the short-term effects after birth could in theory be studied using an interrupted time-series design.

Additionally, our estimates were all drawn from European samples. Because some of the existing literature had found different effects in predominantly non-European non-English speaking populations compared to those observed in English-speaking populations [15–17], our results may not generalise to other populations.

#### 4.3. Strengths and Limitations

This study has several methodological strengths. Firstly, we believe it is the first study to apply MR to explore the effects of the heritable environment in a setting in which gene–environment equivalence is plausible. By doing so, we have been able to leverage the methodological strengths of MR, such as improved robustness to confounding and reverse causation to explore the effect of having children on wellbeing.

Secondly, this is the first applied study to have implemented the MRSamePopTest R package, a novel test of the two-sample MR 'same-population' assumption described in detail elsewhere [58]. We are unaware of other existing tests of this assumption and therefore hope that it will be useful in future applied MR studies.

There are also methodological limitations to this application of two-sample MR. As already noted, we were unable to explore time-sensitive effects. Relatedly, we were forced to assume a linear dose–response relationship for the effect of the number of children on wellbeing. We considered a sensitivity analysis using individual-level data, but we ultimately decided against doing so due to a lack of sufficiently good individual-level data; of the two available data sources, ALSPAC had detailed phenotyping, but on a relatively small number of participants (~2000). Because non-linear MR is less well powered than a linear MR, this sample would therefore be underpowered for this analysis. On the other hand, the UKB, only had a five-level minimal phenotype for happiness. Since poor phenotyping can mask non-linearities, and because happiness may not be the same as wellbeing, the interpretation of any analysis in the UKB would be unclear [65]. MR is generally more robust as a test of the causal null hypothesis than as a method of effect estimation because of many of the complications (such as those described above) of interpreting MR effect estimates. However, this approach to interpreting MR results may be less robust here because our sensitivity analyses implied that our study may be underpowered.

As with many MR studies using genome-wide significant SNPs as instruments, we have not explored the causal mechanism linking the SNPs to having children. However, provided the other Instrumental Variables' assumptions are valid, IV analyses do not require a *causal* association between the instrument and the exposure [66]. Although there was substantive sample overlap between our exposure and the outcome GWAS, we do not expect there to be material bias to our MR results. The one-sample MR two-stage least square estimator is asymptotically equivalent to the two-sample MR Wald ratio, but it is necessarily applied in settings equivalent to 100% sample overlap. If two-sample MR is necessarily biased by sample overlap, then a one-sample MR would not provide valid estimates. Indeed, existing research implies that most two-sample MR estimators perform well even with a 100% sample overlap [67]. Instead, the primary effect of no sample overlap is to ensure that

weak instrument bias attenuates results. However, the F-statistic in our primary analysis is 49, which implies only a 2% bias due to weak instruments. While we cannot guarantee this will attenuate results, it does not seem likely to cause material bias. Finally, we cannot explore more detailed psychological mechanisms, such as differences between expected (numbers of) children and reality, which could explain the observational analyses.

## 5. Conclusions

We conducted a two-sample Mendelian randomisation study to explore the causal effect that children have on parental wellbeing. Contrary to the previous literature, our results do not imply the presence of a negative lifetime effect of having children on wellbeing. Comparing our results to many existing observational studies is complicated by the temporal insensitivity of MR estimates. Future studies could therefore consider using other quasi-experimental methods, such as Interrupted Time Series [68], to explore if the discrepancy between our findings and the observational literature are due to the transient effects of having children, of which MR is either unable to detect or which average out over the life course.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes14030716/s1>, Figure S1: Difference in SNP—outcome associations between the UKB and SSGAC when using a  $5 \times 10^{-8}$  *p*-value threshold; Figure S2: Difference in SNP—outcome associations between the UKB and WFC when using a  $5 \times 10^{-8}$  *p*-value threshold; Figure S3: Difference in SNP—outcome associations between the UKB and SSGAC when using a  $5 \times 10^{-6}$  *p*-value threshold; Figure S4: Forest plot for the SNP-specific Wald ratios for the primary analysis; Figure S5: Funnel plot for the primary analysis; Figure S6: Leave-one-out sensitivity analysis; Table S1: Gene-exposure and gene-outcome associations used in the primary analysis; Table S2: Gene-exposure and gene-outcome associations used in the secondary analysis with an indicative SNP selection threshold; Table S3: Gene-exposure and gene-outcome associations used in the secondary analysis WFC data; Table S4: Full MR results; Table S5: Negative control outcome results; Supplementary Methods. References [37–39,45,46,52,57,69–72] are cited in the Supplementary Materials.

**Author Contributions:** B.W. designed and implemented the study under the supervision of H.M.S. and M.R.M. All authors contributed to the writing of the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** Benjamin Woolf is funded by an Economic and Social Research Council (ESRC) South West Doctoral Training Partnership (SWDTP) 1 + 3 PhD Studentship Award (ES/P000630/1). B.W., H.M.S. and M.R.M. are all members of the Medical Research Council (MRC) Integrative Epidemiology Unit at the University of Bristol (MC\_UU\_00011/7). This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All data that were used in this study are publicly available from the MRC-IEU OpenGWAS platform from the following links: UKB number of children—<https://gwas.mrcieu.ac.uk/datasets/ieu-b-4760/> (accessed on 1 January 2023), SSGAC wellbeing—<https://gwas.mrcieu.ac.uk/datasets/ieu-a-1009/> (accessed on 1 January 2023), and WFC wellbeing—<https://gwas.mrcieu.ac.uk/datasets/ieu-b-4851/> (accessed on 1 January 2023). The R code used in the manuscript is available from <https://doi.org/10.17605/OSF.IO/BTPH9>.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Collins, C.; Glass, J. Effects of Work-Family Policies on Parenthood and Wellbeing. *Handbook of Family Policy*. 30 November 2018. Available online: <https://www.elgaronline.com/view/edcoll/9781784719333/9781784719333.00035.xml> (accessed on 18 March 2022).
2. Herbst, C.M.; Ifcher, J. The increasing happiness of US parents. *Rev. Econ. Househ.* **2016**, *14*, 529–551. [CrossRef]

3. Evenson, R.J.; Simon, R.W. Clarifying the Relationship Between Parenthood and Depression. *J. Health Soc. Behav.* **2005**, *46*, 341–358. [[CrossRef](#)] [[PubMed](#)]
4. McLanahan, S.; Adams, J. Parenthood and Psychological Well-Being. *Annu. Rev. Sociol.* **1987**, *13*, 237–257. [[CrossRef](#)]
5. Hughes, M. Parenthood and Psychological Well-Being Among the Formerly Married: Are Children the Primary Source of Psychological Distress? *J. Fam. Issues* **1989**, *10*, 463–481. [[CrossRef](#)]
6. Keizer, R.; Dykstra, P.A.; Poortman, A.R. The transition to parenthood and well-being: The impact of partner status and work hour transitions. *J. Fam. Psychol.* **2010**, *24*, 429–438. [[CrossRef](#)] [[PubMed](#)]
7. Nomaguchi, K.; Milkie, M.A. Parenthood and Well-Being: A Decade in Review. *J. Marriage Fam.* **2020**, *82*, 198–223. [[CrossRef](#)]
8. Hansen, T. Parenthood and Happiness: A Review of Folk Theories Versus Empirical Evidence. *Soc. Indic. Res.* **2012**, *108*, 29–64. [[CrossRef](#)]
9. Nelson, S.K.; Kushlev, K.; English, T.; Dunn, E.W.; Lyubomirsky, S. In Defense of Parenthood: Children Are Associated With More Joy Than Misery. *Psychol. Sci.* **2013**, *24*, 3–10. [[CrossRef](#)] [[PubMed](#)]
10. Horton, R.S. Parenthood, subjective well-being, and the moderating effects of parent narcissism. *J. Individ. Differ.* **2021**, *42*, 57–63. [[CrossRef](#)]
11. Nelson, S.K.; Kushlev, K.; Lyubomirsky, S. The pains and pleasures of parenting: When, why, and how is parenthood associated with more or less well-being? *Psychol. Bull.* **2014**, *140*, 846–895. [[CrossRef](#)]
12. Nelson-Coffey, S.K.; Killingsworth, M.; Layous, K.; Cole, S.W.; Lyubomirsky, S. Parenthood Is Associated With Greater Well-Being for Fathers Than Mothers. *Pers. Soc. Psychol. Bull.* **2019**, *45*, 1378–1390. [[CrossRef](#)] [[PubMed](#)]
13. Brandel, M.; Melchiorri, E.; Ruini, C. The Dynamics of Eudaimonic Well-Being in the Transition to Parenthood: Differences Between Fathers and Mothers. *J. Fam. Issues* **2018**, *39*, 2572–2589. [[CrossRef](#)]
14. Umberson, D.; Gove, W.R. Parenthood and Psychological Well-Being: Theory, Measurement, and Stage in the Family Life Course. *J. Fam. Issues* **1989**, *10*, 440–462. [[CrossRef](#)]
15. Deaton, A.; Stone, A.A. Evaluative and hedonic wellbeing among those with and without children at home. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 1328–1333. [[CrossRef](#)]
16. Radó, M.K. Tracking the Effects of Parenthood on Subjective Well-Being: Evidence from Hungary. *J. Happiness Stud.* **2020**, *21*, 2069–2094. [[CrossRef](#)]
17. Yu, Q.; Zhang, J.; Zhang, L.; Zhang, Q.; Guo, Y.; Jin, S.; Chen, J. Who Gains More? The Relationship Between Parenthood and Well-Being. *Evol. Psychol.* **2019**, *17*, 1474704919860467. [[CrossRef](#)]
18. Margolis, R.; Myrskylä, M. A Global Perspective on Happiness and Fertility. *Popul. Dev. Rev.* **2011**, *37*, 29–56. [[CrossRef](#)] [[PubMed](#)]
19. Stanca, L. Suffer the little children: Measuring the effects of parenthood on well-being worldwide. *J. Econ. Behav. Organ.* **2012**, *81*, 742–750. [[CrossRef](#)]
20. Novoa, C.; Bustos, C.; Bühring, V.; Oliva, K.; Páez, D.; Vergara-Barra, P.; Cova, F. Subjective Well-Being and Parenthood in Chile. *Int. J. Environ. Res. Public Health* **2021**, *18*, 7408. [[CrossRef](#)] [[PubMed](#)]
21. Neuberger, F.S.; Preisner, K. Parenthood and Quality of Life in Old Age: The Role of Individual Resources, the Welfare State and the Economy. *Soc. Indic. Res.* **2018**, *138*, 353–372. [[CrossRef](#)]
22. Becker, C.; Kirchmaier, I.; Trautmann, S.T. Marriage, parenthood and social network: Subjective well-being and mental health in old age. *PLoS ONE* **2019**, *14*, e0218704. [[CrossRef](#)]
23. Hansen, T.; Slagsvold, B.; Moum, T. Childlessness and Psychological Well-Being in Midlife and Old Age: An Examination of Parental Status Effects Across a Range of Outcomes. *Soc. Indic. Res.* **2009**, *94*, 343–362. [[CrossRef](#)]
24. Munafò, M.R.; Davey Smith, G. Robust research needs many lines of evidence. *Nature* **2018**, *553*, 399–401. [[CrossRef](#)]
25. Davey Smith, G.; Holmes, M.V.; Davies, N.M.; Ebrahim, S. Mendel’s laws, Mendelian randomization and causal inference in observational data: Substantive and nomenclatural issues. *Eur. J. Epidemiol.* **2020**, *35*, 99–111. [[CrossRef](#)]
26. Gage, S.H.; Smith, G.D.; Ware, J.J.; Flint, J.; Munafò, M.R. G = E: What GWAS Can Tell Us about the Environment. *PLoS Genet.* **2016**, *12*, e1005765. [[CrossRef](#)] [[PubMed](#)]
27. Sanderson, E.; Glymour, M.M.; Holmes, M.V.; Kang, H.; Morrison, J.; Munafò, M.R.; Palmer, T.; Schooling, C.M.; Wallace, C.; Zhao, Q.; et al. Mendelian randomization. *Nat. Rev. Methods Prim.* **2022**, *2*, 6. [[CrossRef](#)]
28. Bennett, D.A.; Holmes, M.V. Mendelian randomisation in cardiovascular research: An introduction for clinicians. *Heart* **2017**, *103*, 1400–1407. [[CrossRef](#)] [[PubMed](#)]
29. Gill, D.; Georgakis, M.K.; Walker, V.M.; Schmidt, A.F.; Gkatzionis, A.; Freitag, D.F.; Finan, C.; Hingorani, A.D.; Howson, J.M.M.; Burgess, S.; et al. Mendelian randomization for studying the effects of perturbing drug targets. *Wellcome Open Res.* **2021**. Available online: <https://wellcomeopenresearch.org/articles/6-16> (accessed on 18 March 2022). [[CrossRef](#)]
30. Wootton, R.E.; Jones, H.J.; Sallis, H.M. Mendelian randomisation for psychiatry: How does it work, and what can it tell us? *Mol. Psychiatry* **2021**, *27*, 53–57. [[CrossRef](#)] [[PubMed](#)]
31. McMartin, A.; Conley, D. Commentary: Mendelian randomization and education—Challenges remain. *Int. J. Epidemiol.* **2020**, *49*, 1193–1206. [[CrossRef](#)]
32. Jaffee, S.; Price, T. Gene–environment correlations: A review of the evidence and implications for prevention of mental illness. *Mol. Psychiatry* **2007**, *12*, 432–442. [[CrossRef](#)]
33. Dawkins, R. *The Extended Phenotype: The Long Reach of the Gene*, reprint ed.; OUP Oxford: Oxford, UK, 2016; 496p.

34. Darwin, C. *On the Origin of Species: By Means of Natural Selection or the Preservation of Favoured Races in the Struggle for Life*, 1st ed.; Bynum, W., Ed.; Penguin Classics: London, UK, 2009; 576p.
35. The Selfish Gene (Audio Download): Richard Dawkins, Richard Dawkins, Lalla Ward, Audible Studios: Amazon.co.uk: Audible Books & Originals. Available online: [https://www.amazon.co.uk/The-Selfish-Gene/dp/B004UAZL5U/ref=sr\\_1\\_1?keywords=the+selfish+gene&qid=1647643192&prefix=the+selfish+g%2Caps%2C119&sr=8-1](https://www.amazon.co.uk/The-Selfish-Gene/dp/B004UAZL5U/ref=sr_1_1?keywords=the+selfish+gene&qid=1647643192&prefix=the+selfish+g%2Caps%2C119&sr=8-1) (accessed on 18 March 2022).
36. Collins, R. What makes UK Biobank special. *Lancet* **2012**, *379*, 1173–1174. [[CrossRef](#)] [[PubMed](#)]
37. Elsworth, B.; Lyon, M.; Alexander, T.; Liu, Y.; Matthews, P.; Hallett, J.; Bates, P.; Palmer, T.; Haberland, V.; Smith, G.D.; et al. The MRC IEU OpenGWAS data infrastructure. *bioRxiv* **2020**. Available online: <https://www.biorxiv.org/content/10.1101/2020.08.10.244293v1> (accessed on 30 March 2022).
38. Okbay, A.; Baselmans, B.M.L.; De Neve, J.E.; Turley, P.; Nivard, M.G.; Fontana, M.A.; Meddens, S.F.W.; Linnér, R.K.; Rietveld, C.A.; Derringer, J.; et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* **2016**, *48*, 624–633. [[CrossRef](#)]
39. Howe, L.J.; Nivard, M.G.; Morris, T.T.; Hansen, A.F.; Rasheed, H.; Cho, Y.; Chittoor, G.; Lind, P.A.; Palviainen, T.; van der Zee, M.D.; et al. Within-sibship GWAS improve estimates of direct genetic effects. *bioRxiv* **2021**. Available online: <https://www.biorxiv.org/content/10.1101/2021.03.05.433935v1> (accessed on 12 March 2022).
40. Davies, N.M.; Howe, L.J.; Brumpton, B.; Havdahl, A.; Evans, D.M.; Davey Smith, G. Within family Mendelian randomization studies. *Hum. Mol. Genet.* **2019**, *28*, R170–R179. [[CrossRef](#)]
41. Bigdeli, T.B.; Lee, D.; Riley, B.P.; Vladimirov, V.; Fanous, A.H.; Kendler, K.S.; Bacanu, S.A. FIQT: A simple, powerful method to accurately estimate effect sizes in genome scans. *bioRxiv* **2015**, 019299. Available online: <https://www.biorxiv.org/content/10.1101/019299v1> (accessed on 12 March 2022).
42. Slob, E.A.W.; Burgess, S. A comparison of robust Mendelian randomization methods using summary data. *Genet. Epidemiol.* **2020**, *44*, 313–329. [[CrossRef](#)] [[PubMed](#)]
43. Bowden, J.; Del Greco, M.F.; Minelli, C.; Davey Smith, G.; Sheehan, N.A.; Thompson, J.R. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: The role of the I<sup>2</sup> statistic. *Int. J. Epidemiol.* **2016**, *45*, 1961–1974.
44. Bowden, J.; Davey Smith, G.; Burgess, S. Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **2015**, *44*, 512–525. [[CrossRef](#)] [[PubMed](#)]
45. Zhao, Q.; Wang, J.; Hemani, G.; Bowden, J.; Small, D.S. Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. *Ann. Stat.* **2020**, *48*, 1742–1769. Available online: <http://arxiv.org/abs/1801.09652> (accessed on 17 April 2019). [[CrossRef](#)]
46. Verbanck, M.; Chen, C.Y.; Neale, B.; Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* **2018**, *50*, 693–698. [[CrossRef](#)] [[PubMed](#)]
47. Hartwig, F.P.; Wang, L.; Smith, G.D.; Davies, N.M. Average causal effect estimation via instrumental variables: The no simultaneous heterogeneity assumption. *arXiv* **2021**. Available online: <http://arxiv.org/abs/2010.10017> (accessed on 28 June 2022). [[CrossRef](#)] [[PubMed](#)]
48. Burgess, S.; Davey Smith, G.; Davies, N.M.; Dudbridge, F.; Gill, D.; Glymour, M.M.; Hartwig, F.P.; Holmes, M.V.; Minelli, C.; Relton, C.L.; et al. Guidelines for performing Mendelian randomization investigations. *Wellcome Open Res.* **2020**, *4*, 186. [[CrossRef](#)] [[PubMed](#)]
49. Davies, N.M.; Holmes, M.V.; Smith, G.D. Reading Mendelian randomisation studies: A guide, glossary, and checklist for clinicians. *BMJ* **2018**, *362*, k601. [[CrossRef](#)] [[PubMed](#)]
50. Woolf, B.; Di Cara, N.; Moreno-Stokoe, C.; Skrivankova, V.; Drax, K.; Higgins, J.P.T.; Hemani, G.; Munafò, M.R.; Smith, G.D.; Yarmolinsky, J.; et al. Investigating the transparency of reporting in two-sample summary data Mendelian randomization studies using the MR-Base platform. *Int. J. Epidemiol.* **2022**, *51*, dyac074. [[CrossRef](#)] [[PubMed](#)]
51. Burgess, S.; Davies, N.M.; Thompson, S.G. Instrumental Variable Analysis with a Nonlinear Exposure–Outcome Relationship. *Epidemiology* **2014**, *25*, 877–885. [[CrossRef](#)]
52. Loh, P.-R.; Tucker, G.; Bulik-Sullivan, B.K.; Vilhjálmsson, B.J.; Finucane, H.K.; Salem, R.M.; Chasman, D.; Ridker, P.M.; Neale, B.M.; Berger, B.; et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **2015**, *47*, 284–290. [[CrossRef](#)]
53. Bulik-Sullivan, B.K.; Loh, P.R.; Finucane, H.K.; Ripke, S.; Yang, J.; Schizophrenia Working Group of the Psychiatric Genomics Consortium; Patterson, N.; Daly, M.J.; Price, A.L.; Neale, B.M. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **2015**, *47*, 291–295. [[CrossRef](#)]
54. Lesko, C.R.; Buchanan, A.L.; Westreich, D.; Edwards, J.K.; Hudgens, M.G.; Cole, S.R. Generalizing study results: A potential outcomes perspective. *Epidemiology* **2017**, *28*, 553–561. [[CrossRef](#)]
55. Dahabreh, I.J.; Hernán, M.A. Extending inferences from a randomized trial to a target population. *Eur. J. Epidemiol.* **2019**, *34*, 719–722. [[CrossRef](#)] [[PubMed](#)]
56. Sanderson, E.; Richardson, T.G.; Hemani, G.; Davey Smith, G. The use of negative control outcomes in Mendelian randomization to detect potential population stratification. *Int. J. Epidemiol.* **2021**, *50*, 1350–1361. [[CrossRef](#)] [[PubMed](#)]

57. Skrivankova, V.W.; Richmond, R.C.; Woolf, B.A.R.; Davies, N.M.; A Swanson, S.; VanderWeele, T.J.; Timpson, N.J.; Higgins, J.P.T.; Dimou, N.; Langenberg, C.; et al. Strengthening the reporting of observational studies in epidemiology using mendelian randomisation (STROBE-MR): Explanation and elaboration. *BMJ* **2021**, *375*, n2233. [[CrossRef](#)] [[PubMed](#)]
58. Woolf, B.; Sallis, H.; Munafo, M.; Gill, D. MRSamePopTest: Introducing a simple falsification test for the Two-Sample Mendelian randomisation ‘same population’ assumption. *OSF Prepr.* **2022**. Available online: <https://osf.io/gvt87/> (accessed on 21 June 2022).
59. Staley, J.R.; Blackshaw, J.; Kamat, M.A.; Ellis, S.; Surendran, P.; Sun, B.B.; Paul, D.S.; Freitag, D.; Burgess, S.; Danesh, J.; et al. PhenoScanner: A database of human genotype-phenotype associations. *Bioinformatics* **2016**, *32*, 3207–3209. [[CrossRef](#)]
60. Giannelis, A.; Palmos, A.; Hagenaaars, S.P.; Breen, G.; Lewis, C.M.; Mutz, J. Examining the association between family status and depression in the UK Biobank. *J. Affect. Disord.* **2021**, *279*, 585–598. [[CrossRef](#)] [[PubMed](#)]
61. Fabbri, C.; Mutz, J.; Lewis, C.M.; Serretti, A. Stratification of individuals with lifetime depression and low wellbeing in the UK Biobank. *J. Affect. Disord.* **2022**, *314*, 281–292. [[CrossRef](#)]
62. Labrecque, J.A.; Swanson, S.A. Interpretation and Potential Biases of Mendelian Randomization Estimates With Time-Varying Exposures. *Am. J. Epidemiol.* **2019**, *188*, 231–238. [[CrossRef](#)]
63. Sanderson, E.; Richardson, T.G.; Morris, T.T.; Tilling, K.; Smith, G.D. Estimation of causal effects of a time-varying exposure at multiple time points through Multivariable Mendelian randomization. *medRxiv* **2022**. Available online: <https://www.medrxiv.org/content/10.1101/2022.01.04.22268740v1> (accessed on 11 May 2022). [[CrossRef](#)]
64. Morris, T.T.; Heron, J.; Sanderson, E.C.M.; Davey Smith, G.; Didelez, V.; Tilling, K. Interpretation of Mendelian randomization using a single measure of an exposure that varies over time. *Int. J. Epidemiol.* **2022**, *51*, dyac136. [[CrossRef](#)] [[PubMed](#)]
65. van Smeden, M.; Lash, T.L.; Groenwold, R.H.H. Reflection on modern methods: Five myths about measurement error in epidemiological research. *Int. J. Epidemiol.* **2020**, *49*, 338–347. [[CrossRef](#)] [[PubMed](#)]
66. Angrist, J.D.; Pischke, J.S. *Mostly Harmless Econometrics: An Empiricist’s Companion*, illustrated ed.; Princeton University Press: Princeton, NJ, USA, 2009; 392p.
67. Minelli, C.; Del Greco, M.F.; van der Plaats, D.A.; Bowden, J.; Sheehan, N.A.; Thompson, J. The use of two-sample methods for Mendelian randomization analyses on single large datasets. *Int. J. Epidemiol.* **2021**, *50*, 1651–1659. [[CrossRef](#)] [[PubMed](#)]
68. Ferron, J.; Rendina-Gobioff, G. Interrupted Time Series Design. In *Encyclopedia of Statistics in Behavioral Science*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2005; Available online: <https://onlinelibrary.wiley.com/doi/abs/10.1002/0470013192.bsa312> (accessed on 11 May 2022).
69. Ruth Mitchell, E. MRC IEU UK Biobank GWAS Pipeline Version 2. 2019. Available online: <https://data.bris.ac.uk/data/dataset/pnoat8cxo0u52p6ynfaekeigi> (accessed on 12 March 2022).
70. Hemani, G.; Zheng, J.; Elsworth, B.; Wade, K.H.; Haberland, V.; Baird, D.; Laurin, C.; Burgess, S.; Bowden, J.; Langdon, R.; et al. The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **2018**, *7*, e34408. [[CrossRef](#)] [[PubMed](#)]
71. Schwarzer, G.; Carpenter, J.R.; Rücker, G. *Meta-Analysis with R*; Springer International Publishing: Cham, Switzerland, 2015; Available online: <http://link.springer.com/10.1007/978-3-319-21416-0> (accessed on 14 April 2022).
72. Textor, J.; van der Zander, B.; Gilthorpe, M.S.; Liškiewicz, M.; Ellison, G.T. Robust causal inference using directed acyclic graphs: The R package ‘dagitty’. *Int. J. Epidemiol.* **2016**, *45*, 1887–1894. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.