Referee Report for PNAS

Manuscript 2021 – 02781: "Natural selection in contemporaneous humans is liked to income and substitution effects"

This submission contributes to the literature on natural selection in humans by documenting associations between fertility (number of children) and polygenic scores for 33 different traits in the UK Biobank. The main results reveal statistically significant associations between several polygenic scores and number of children. To highlight a few specific results - higher values of the scores for ADHD, Smoking behavior, BMI, and Major Depressive Disorder tend to be positive associated with number of children, while higher values of the scores for educational attainment are found to be negatively associated with number of children. The authors conduct a number of exercises that split the sample by important observables or outcomes. They find that the scores that predict number of children tend to have larger associations with fertility among individuals with lower levels of income, those with a lower age at first live birth (AFLB), those with more sexual partners, and those not living with a partner. Most of the largest associations appear to be mediated by educational attainment and income.

The authors interpret the results through the lens of economic theories of fertility. In particular, in a Becker and Tomes (1976) framework, one would expect offsetting income and substitution effects in the relationship between household income and number of children. More economic resources allow households to afford more children (income effect). At the same time, individuals with higher incomes tend to have higher opportunity costs of time spent raising kids, reducing fertility (substitution effect). The authors argue that their empirical results can be interpreted as providing evidence that the substitution effect dominates the income effect overall, and especially at low levels of the income distribution. The positive association between scores like the educational attainment score and fertility, conditioning on AFLB, is interpreted as providing evidence of the operation of an income effect.

I think that the descriptive results in this paper are very interesting. It is useful to see patterns of association between various polygenic scores and fertility. I also applaud the attempt to connect these patterns to economic theory. However, I have several major reservations about how that part of the analysis is conducted. Overall, I think that the paper needs to provide a clearer statement about either (1) what more we learn about income and substitution effects relative to existing studies that directly examine the relationship between income and fertility, or (2) what we learn about genetic links to fertility by using economic theory to interpret relationships between polygenic scores and fertility.

Comments:

- **Collider Bias / Endogeneity of Controls** A large part of the paper is devoted to showing how associations between various polygenic scores and fertility are (1) different across various subpopulations defined by education, income, age at first live birth (AFLB), etc, and (2) altered when controlling for these variables in regressions. These differences are interpreted as providing evidence on the strength of income and substitution effects in the relationship between household resources and fertility. A problem here is that these control variables are all either likely to be affected by the polygenic scores (e.g. the polygenic score for education), or themselves reflect a fertility outcome (e.g. AFLB). This means that controlling for these variables or splitting the sample by their values can introduce collider bias that makes the results in this paper difficulty to interpret. For example, consider splitting the sample by educational attainment (EA) and looking at associations between the EA polygenic score and fertility. Those individuals with low EA polygenic scores who have high levels of observed EA will be non-randomly selected on other characteristics that promote educational attainment. If these other characteristics are negatively associated with fertility, this would tend to bias a true negative relationship between the EA polygenic score and fertility towards zero. This would make it difficult to interpret the lack of an association between the EA polygenic score and fertility (given high education) as

evidence of offsetting income and substitution effects in this group.

Splitting the sample by AFLB presents additional challenges because AFLB is itself one aspect of the fertility outcome. AFLB is not (just) a possible mediating mechanism, but it is jointly determined with the rest of lifetime fertility. Households choosing to have more children will tend to have a lower AFLB, so examining the relationship between any polygenic score and fertility *conditional on AFLB* is very difficult to interpret.

- **Interpretation of the Theory** I think the authors need to substantially revisit how they are connecting the empirical results here with economic theories of fertility. As a general comment, I think the authors should offer a more formal exposition of the theory. I have three specific comments related to the theory:

    1. **Role of Age at First Live Birth:** The authors write the following about age at first live birth on p. 11: "If so, then people with less human capital will leave education earlier and have their first child earlier. This is a pure substitution effect: the income effect plays no role in the decision to leave education, since people are not yet earning wages." I think there are several things that the authors need to address about this statement. First, I do not understand exactly what it means for the income effect to play no role in the decision to leave school. Are the authors trying to say that relationships between polygenic scores (linked to human capital) and AFLB reflect only a pure substitution effect? If this is not what is meant, then I do not understand what the authors are trying to say. If this is what is meant, then the authors need to provide a clearer theoretical argument for why this is true. The idea that future wages do not matter for current fertility is clearly incorrect when there are asset markets that allow individuals to borrow and save. The authors mention this in footnote 4, but this needs to be much more clearly stated and explained in the text since it is a core assumption here. Furthermore, even without saving / borrowing, I still think future wages will

matter because each child generates a stream of costs throughout several future periods. The price of having a child now is not paid only in the present - having a child now commits the household to a sequence of future costs which will have to be paid out of future earnings. Even in a model with no savings and borrowing, expectations of future wages will still affect fertility decisions earlier in life.

2. **Dynamic Theory:** I may be repeating a bit from the previous point, but any prediction involving AFLB in a serious way needs to be grounded in an explicitly dynamic theory of fertility. As a static model, the Becker and Tomes framework does not provide guidance on this. The authors should have a discussion about empirical predictions (especially for AFLB) that is grounded in dynamic theory. Adda, Dustmann and Stevens (2017, JPE, "The Career Costs of Children) provide a recent example of a dynamic model of fertility and labor market outcomes that may be useful here.

3. **Male v.s. Female Polygenic Scores:** In the Becker and Tomes setup, there is not an explicit distinction between choices made by male and female household members. In a more complete model, one would imagine polygenic scores for men and women affecting fertility in different ways. Empirically men are less likely to stay at home or adopt primary caregiving roles, either because their wages are higher, or because of societal norms / expectations, or because they have higher costs to supplying childcare at home, etc. This might suggest under some assumptions, that the relationships between fertility and male scores would be more directly related to income effects, while the relationships between fertility and female scores would pick up more of a mix of income and substitution effects. This kind of a distinction might provide a clearer test of the relative strength of the income and substitution effects. It is noteworthy that, in Appendix Figure 8, there is no statistically significant difference in the coefficients on the EA polygenic scores across men or women. This would seem to be informative of the

kinds of mechanisms that are at play here. I would think that, if the substitution effect were explaining the relationship between the EA polygenic scores and fertility, that one would see a substantially larger negative relationship between female polygenic scores and fertility as opposed to male polygenic scores and fertility - but that is not what seems present in the data. Of course, this raises the larger question about assortative mating, since an individual's polygenic scores are correlated with their spouse's scores. This suggests that it might be more appropriate to consider the household as the unit of analysis and enter the polygenic scores of both spouses separately into the empirical specification.

- **Testing Alternate Theories** The authors try to test the "economic theory" against alternate theories explaining links between polygenic scores and fertility (they mention externalizing behavior, risk-seeking, high time discount rates, etc). One of the tests of the economic theory is that many of the scores become insignificant predictors of fertility once controlling for (initial) earnings and education. I think this test of the theory is problematic, since the associations between earnings / education and fertility might be coming about precisely because they are correlated with unobserved traits like time preference. It could be the case that these polygenic scores exclusively work through time preference (or other mechanisms), and that individuals with high discount rates are both less likely to acquire human capital and more likely to have more children. In this case, controlling for earnings and education would appear to account for the score-fertility relationships, but this might only reflect the fact that earnings and education are endogenously determined as functions of time preference, for example. I would also note here that I find the language that the authors use to be misleading. The authors suggest that "the economic theory" consists entirely of predictions about how household resources are related to the number of children. However, economic theories of fertility also predict that time preferences, risk preferences, and other primitive parameters should determine fertility outcomes since these parameters affect the

decision-making process. It is not clear that one should consider a theory that links time preference to fertility any less of an "economic theory."

- **Testing Alternate Theories - Risk Measure** The risk taking measure in UK Biobank field 2040 is a binary response to the question "Would you describe yourself as someone who takes risks?". I think that many readers will doubt whether this sufficiently captures variation in risk preferences, let alone impulsiveness, externalizing behavior, time preferences, and other primitive parameters that might generate the joint relationships between the polygenic scores, income / earnings, and fertility. At the very least, the serious limitations of this measure should be discussed, and I would hesitate to describe the inclusion of this variation as a control as a test of the "alternative theory."

- **Polygenic score for fertility** I am surprised that the authors did not make use of polygenic scores for fertility based on the GWAS results in Barban et al (2016). It would be interesting to understand the degree of correlation between polygenic scores for fertility and the 33 scores examined here. Moreover, it would be useful to understand the relative predictive power of the polygenic score for fertility v.s. the other polygenic scores. For example, it could be the case that the 33 scores used in this analysis seem to be linked to fertility through education / earnings, but they might still account for a small amount of the relationship between the *fertility polygenic score* and observed fertility. This is important because the paper is written in a way that suggests that these 33 polygenic scores are sufficient to understand patterns of natural selection. If these scores do not capture much of the genetic links represented by the fertility scores, then this claim would seem to be quite incomplete.

- **Reverse Causality** Related to the previous point, I think many readers will question whether there might be reverse causality in play here. Specifically, early child-bearing can disrupt educational and career trajectories, generating a causal effect of fertility

on education and labor market outcomes. The genetic variants detected by a GWAS for education, are likely to operate through many different indirect channels. It could be the case that some of the variants that are weighted in the EA polygenic score represent markers that causally affect fertility, which in turn affects educational attainment. Understanding the correlation between the polygenic scores studies here and the polygenic score for fertility would provide some context for thinking about this, and possibility for judging how much the results might be driven by this kind of reverse causality.

- **Counterfactuals / Contribution** I think the authors could do a better job of establishing the contribution of this analysis above and beyond the existing literature on fertility. Existing literatures in economics and related fields estimate the relationship between income and fertility, as well as the intergenerational persistence of income and education (among other outcomes). It would be helpful if the authors could provide an argument for what the study of genetic variation adds to our understanding of selection into higher rates of fertility and what implications this might have for future income and health inequality.

- **Emphasizing Main Results** I find it a bit jarring that the authors immediately jump into an analysis of how the coefficients on the polygenic scores differ by subgroups. Figure 2 presents the first regression results, and right away these reflect sample splits. I think it would be more informative for the reader to see a plot that simply shows the baseline coefficients that one gets in the entire sample.

- **Language** I think that the authors should be a bit more careful with the kind of language that is used in the discussion section to describe the traits that are being selected for (or against). Much of the results in the paper are centered around genetic variation linked to educational attainment. The authors write that "many of the phenotypes under positive selection are linked to disease risk, or are what many people

would see as undesirable to have." I worry that this kind of language could be easily misread as implying that natural selection is favoring "undesirable" types of people. I am confident that this is NOT the intention of the authors. However, as someone who has followed this literature closely, I am acutely aware of the (justified) sensitivities that many people have when discussing genetic variation and social science outcomes. I just want to suggest that the authors revise the language used here to prevent any uncharitable interpretations. I think that this also applies to the discussion of "biocosmic pessimism" and "dysgenic selection" among alien species. This connection seems wildly speculative and unnecessary for the current analysis. I realize that the authors are suggesting that their results provide some nuance to these kinds of debates, but I think a prudent strategy would be to avoid language like this.