EPC 2014, Budapest, Hungary, 25-28 June 2014

Do Genetic Markers for Infertility Problems Predict Childlessness and Completed Fertility?

Jornt Mandemakers[1], Nicola Barban[1], Melinda Mills[1], Harold Snieder[2]

[1]University of Groningen, The Netherlands. Department of Sociology
[2]University of Groningen, University medical center Groningen, The Netheralands. Department of Epidemiology

Corresponding author: j.j.mandemakers@rug.nl

Draft prepared for EPC 2014. Please do not quote.

ABSTRACT

We present a novel approach to examine the influence of biological limits to reproduction. We use a set of validated genetic markers from published GWAS studies on phenotypes related to infertility (endometriosis and (early) menopause) in order to create polygenic infertility problem risk scores. We hypothesize that women carrying more 'infertility risk' alleles are at increased risk of childlessness and decreased completed fertility. Second, we hypothesize that women from later cohorts and more educated women will be more vulnerable to genetically endowed infertility problems, as they may postpone childbearing to a greater extent. Preliminary analyses using the Dutch LifeLines cohort show that a higher predicted genetic risk increases the likelihood of childlessness and lowers completed fertility, but only for lower educated women. Contrary to the postponement hypothesis, higher educated women are less affected by predicted genetic risk scores. We interpret these educational differences as protective effects because higher educated women may have better health, less stress, healthier lifestyle, and more access to health care. Future versions of this paper will replicate these analyses using the TwinsUK and HRS cohorts.

Do Genetic Markers for Infertility Problems Predict Childlessness and Completed Fertility?

The average age at which mothers have their first child has dramatically increased in the last decades in developed countries (Billari, Kohler et al. 2007). Postponing childbearing to these more advanced ages, however, comes with increased risks as women's ability to have children declines with age. Mirowsky (2005) eloquently describes the postponement of childbearing by women as a trade-off between organic and social resources. Women postpone childbearing to attain more social resources, but if they delay their decision for too long, they run higher risks of infertility problems. And consequently a higher chance of involuntary childlessness or a lower than desired number of children. The female biological ability to conceive a child (fecundity) is at its peak around age 25 and starts to decline fast by age 30, and decrease quickly to very low levels after age 35. However, we know very little of the role that biological differences in reproductive ability play in observed reproductive behavior. The widespread use of Assisted Reproductive Technologies, such as IVF, by many people coupled with natural childbearing at advanced ages, shows that biological differences are probably quite important.

Recent advances in genetics may help to shed light on this question. In this paper we focus on the reproduction of women, as more is known about the genetic architecture of female reproductive ability. We present an attempt to integrate new genetic insights into the study of female reproductive behavior. Variation in the ability to conceive and in the age-related decline of reproductive ability may in part be due to genetic differences, as recent studies have identified many genetic variants associated with diseases related to male and female reproductive ability (Montgomery, Zondervan et al. 2013; Zhao and Chen 2013).

We hypothesize that there are observable genetic variants associated with diseases that cause infertility problems and that information about such genetic variants can be used to predict childlessness and completed fertility by constructing a genetic risk score. Second, we hypothesize that genetic differences will be most predictive for people who are most at risk of infertility problems, namely the people that postpone childbearing to later ages. To be precise, we hypothesize that women born in later birth cohorts and more educated women will be more vulnerable to genetically endowed infertility problems, as they may postpone childbearing to a greater extent.

A major problem in this type of research (candidate gene studies) has been the lack of replicability of results (Duncan and Keller 2011; Hewitt 2012). To tackle this problem we base the genetic risk score for infertility problems on results of large GWAS (Genome-Wide Association Studies) findings. And we will try to replicate this analysis across several samples. For this article we aim to replicate the analysis across three different datasets: the Dutch LifeLines Biobank and Cohort, TwinsUK registry and the US-based Health and Retirement Survey. For the preliminary analysis carried out for this version of this paper we currently only include the preliminary analyses for the LifeLines cohort.

We create two infertility problems risk scores based on recent GWAS findings for two of the three most important causes of female infertility, namely early menopause and endometriosis. We use GWAS that discovered common genetic variants associated with (early) menopause (Stolk, Perry et al. 2012; Perry, Corre et al. 2013) and endometriosis (Painter, Anderson et al. 2011; Nyholt, Low et al. 2012; Albertsen, Chettier et al. 2013). For polycystic ovary syndrome, which is another important cause of female infertility, no GWAS in sample of European ancestry has been published as of yet.

Methods

Fertility outcomes

We focus on childlessness (see table 2) and number of children ever born at age 45 (see table 3) as infertility problems will most likely affect the quantum of fertility. Furthermore, possible tempo effects are more complicated to investigate. To account for historical changes in fertility, all analyses allow for a

curvilinear trend for year of birth of the respondent. All analyses also control for educational level, as this is an important indicator of fertility behavior. In appendix II we show that the polygenic score for menopause predicts self-reported menopause and that the self-reported menopause in post-menopausal women is related to their achieved fertility. Taking these results together would suggest that the menopause polygenic score predicts fertility.

Sample - LifeLines

The Dutch LifeLines Biobank and Cohort is a three-generation longitudinal family design of 165,000 individuals from the Northern provinces of the Netherlands(Stolk, Rosmalen et al. 2008). About 7,800 unrelated women have been genotyped so far. By 2013, two waves will be collected and available, with additional waves each year. LifeLines has the advantage of a large sample, ability to separate non-genetic and genetic familial transmission, single and multiple SNPs and direct haplotype assessment. Currently we analyze only the first wave of data, the final version of the paper will incorporate information form the second wave.

For these analyses we selected all genotyped women born after 1930, who were at least 45 at last observation in wave 1 (born in 1965 at the latest). The few women (32 cases) born before 1930 (1920-1929) showed a very high proportion of childlessness (42%), so we decided to exclude these atypical women, as this may be due to selection of the genotyped sample or problems of recall. After listwise deletion the sample is 4,508.

As discussed before, we will replicate the analysis using the TwinsUK registry and the Health and Retirement Survey.

Variables

See table 1 for the descriptive statistics of all the variables in the analysis. Most variables do not need comment, so we only discuss the construction of the polygenic scores.

Menopause & endometriosis polygenic scores

The polygenic scores for menopause was constructed by the count of the number of alleles that are associated with increased age of menopause. We used the 17 SNPs that showed independent association with menopausal age, as reported in the largest and most recent meta-GWAS of menopausal age in people of European ancestry (Stolk, Perry et al. 2012). A recent follow-up GWAS for early menopause (before age 40) (Perry, Corre et al. 2013) showed that these previous 17 reported independent hits for overall menopause age show associations with early menopause in the expected direction. However, not all of the 17 SNPs was replicated analyses of early menopause using the strict Bonferonni genome-wide corrected p-values(Perry, Corre et al. 2013). As we wish to build a polygenic score using as much information as possible we use all of the 17 SNPs (Hewitt 2012). Please see table A1 for information on these SNPs.

The polygenic endometriosis score was created by the count of endometriosis risk enhancing alleles (see table A2), as reported in the largest GWAS of endometriosis in a sample of European ancestry (Albertsen, Chettier et al. 2013). A number of other GWAS studies report associations with endometriosis (Painter, Anderson et al. 2011; Nyholt, Low et al. 2012) and the findings are quite similar, for the present analyses we decided to keep it simple and limit the score to SNPs reported in just one study. Note that we again use all the reported SNPs not just the genome-wide significant hits.

Preliminary results

Tables 2 (childlessness) and 3 (completed fertility) examine the predictive value of the polygenic scores. Menopause is examined in panel A, endometriosis in panel B. The menopause count is expected to be positively related with fertility outcomes, as a later age at menopause extends the reproductive window (see Appendix II for substantiating evidence). The endometriosis risk allele count is expected to be negatively related to fertility, as endometriosis is related to reduced fertility. Furthermore, we expect that

the two interactions of the polygenic scores with birth year and with educational level amplify the effect of the polygenic scores, as we expect that especially higher educated women and later born women tend to postpone childbearing and would thus be more vulnerable to a genetically endowed risk.

Childlessness

We first turn to childlessness in table 2. The second column of table 2 shows the baseline model. As expected, there is a curvilinear pattern in childlessness; pre-WWII cohorts and more recent cohorts experience more childlessness than generation in-between (baby-boomers). Education does not significantly influence childlessness. In the second model (column 3) we add genetic risk scores. The model with the polygenic menopause score is shown in panel A and in a separate model the polygenic endometriosis score in panel B. These models show that having a higher genetic risk for later menopause does not significantly decrease the odds of childlessness as we expected. For endometriosis there is some tentative evidence that having a higher genetic risk seems to decrease the odds of childlessness. The effect is negative (we expected a positive sign), but only significant at the 10% level. So contrary to expectations; for each endometriosis risk allele the odds of childlessness decrease by .011.

The third model (column 4) adds an interaction with birth year to each of the two models to examine the second hypothesis. Namely, whether a genetic infertility risk may be a stronger predictor in more recent cohorts due to postponement of childbearing. That does not seem to be the case, as the interactions are not significant. The fourth model (column 5) adds interactions with educational level to examine the same hypothesis, as especially more highly educated women may postpone childbearing. Contrary to expectations, we find a positive interaction with educational level (albeit only significant at the 10% level), so more highly educated women with a lower genetic risk (later predicted menopause is 'good') are more likely to become childless. For the endometriosis polygenic score we find a significant negative interaction, which is again opposite of what we expected to find. More highly educated women who have more endometriosis risk alleles (a higher genetic risk) are less likely to remain childless. Apparently education has protective effects. Including the interactions with educational level shows that the main effect of the polygenic scores becomes (marginally) significant. This indicates that only for lower educated women the likelihood of childlessness is affected by their genetic endowments. The fifth model (column 6) estimates the two interactions simultaneously, but the interactions effects hardly differ from the previous model.

Completed fertility

The results for completed fertility are similar to those for childlessness. The main difference is that only the endometriosis risk score matters for completed fertility; the menopause risk score does not predict completed fertility at all.

Discussion

In this paper we presented a novel approach to examine biological limits to reproduction. We used a set of validated genetic markers from published GWAS studies (for endometriosis and (early) menopause) to create genetic infertility risk scores. We hypothesized that a genetic risk for infertility increases the likelihood of childlessness and lowers completed fertility. In a second step, we hypothesized that women from later cohorts and more educated women will be more vulnerable to their genetic infertility risk, as they may have postponed childbearing to more risky ages. We examined these hypotheses in the large Dutch LifeLines cohort of about 5,000 genotyped women of post reproductive age.

Preliminary and tentative results using only the LifeLines cohort show that a higher genetic infertility risk increases the likelihood of childlessness and lowers completed fertility, but only for lower educated women. Unexpectedly, highly educated women were less affected by their predict genetic risk for infertility problems. We interpret these findings as protective effects because of increased health, less stress, better lifestyle (less smoking), and better access to health care (infertility treatment) for higher

educated women. Health, stress and lifestyle factors are shared predictors for (early)menopause, endometriosis and fertility behavior, so future analyses will try to include measures for these factors to try to disentangle this puzzle.

A further striking finding is that the results using either a polygenic score for menopause or for endometriosis were quite similar, even though the two polygenic scores were constructed in very different ways. The SNPs to construct the polygenic scores were derived from GWAS that had non-overlapping samples, and the menopause score was based on only relatively few (17) replicated independent hits, whereas for endometriosis we used all (about 100) top hits.

It is important to note that we found effects using the polygenic risk scores, even though the road from SNP to childlessness and completed fertility is a very long one. First, the traits related to infertility we examined are only partly heritable, as opposed to fully heritable Mendellian traits. Twin and family studies show that early menopause and endometriosis are both about 50% heritable. There rest is environmental. Second, the link between infertility and childlessness/completed fertility crucially depends on reproductive choice. Early menopause probably only affects women who postpone childbearing to later ages. Third, the present analysis only in about 40-60% of couples that have infertility problems the problems are due to the infertility problems of the female. Fourth, the polygenic scores we used are based on GWAS findings, but GWAS studies only capture a portion of the heritability that is reported in twin studies. For instance, heritability is about 50% for menopause, but all the SNPs on a chip capture only 22% of the phenotypic variance in menopause. GWAS studies can find common genetic variants that are the actual casual variants or SNPs that are in linkage disequilibrium with a causal variant. So, polygenic scores based on GWAS studies will only give an approximation of the true unknown genetic risk. All in all, despite the discussed difficulties, the presented results suggest that GWAS based genetic risk scores are a viable strategy to investigate the role of biological limits in reproductive choice.

Finally, it is crucial to stress that these analyses need to be replicated in independent samples, as previous candidate gene studies have often failed to replicate (Duncan and Keller 2011; Hewitt 2012). Pending replication the tentative results presented here should be treated with extreme caution. In future versions of this paper, we will replicate the analysis also using the TwinsUK registry and the Health and Retirement Survey.

Future analyses

We plan to carry out more comprehensive analyses and also discuss the assumptions of using genetic markers for prediction in more detail. Furthermore, the LifeLines fertility variables are currently limited to the first 6 children (which may make it harder to detect menopause effects) and the analyses are based on the first wave of data, i.e. this mainly excludes younger cohorts who postponed the most. A second wave will come available by the end of 2013, so we plan to incorporate this wave and probably significantly increase the effective sample size.

Most importantly, we aim to replicate the analyses of LifeLines in at least two other large datasets. Currently we have access to the TwinsUK and are working on similar analyses. As noted above, we plan to use the HRS as well. As we are also carrying out a large meta-analytic GWAS on reproductive behavior (see http://www.ssgac.org/Phenotypes.php) ourselves, we will likely be able to recruit more cohorts for replication.

Future analyses will examine sensitivity of the results to the construction of the polygenic scores (counts versus weighted counts, excluding imputed SNPs, HWE, LD-structure, call-rate, etc., etc.) and discuss in detail underlying assumptions. Moreover, we plan to carry out GCTA analyses using these sets of SNPs and also to examine the age at last birth. Analysis -not shown due to space restrictions- show that the polygenic endometriosis score predicts age at last birth (negative relationship), but the menopause predictor does not. In addition, we need to contact the authors of published GWAS studies on infertility related diseases to see whether they are willing to share top 100 SNPs (e.g. these are not in the public domain for menopause, but are published online for endometriosis). Moreover, we plan to extend these analyses as novel loci are discovered for infertility related diseases in people of European ancestry (e.g. a GWAS for polycystic ovary syndrome, or larger GWAS for menopause, etc.).

Tables & figures

Table 1. Descriptives of variables in LifeLines (female genotyped subsample).

	mean	s.d.	min	max	N
birth year	1954.15	7.86	1930	1965	4698
age at censoring	55.04	8.17	45	80.7	4698
female	1	-	1	1	4698
educational level (years)	13.44	2.31	8	18	4508
age at menopause	47.86	6.06	30	68	2714
number of children at age 45	2.29	1.20	0	6	4698
childless by 45	.11	-	0	1	4698
age at last child birth	30.68	4.44	16.8	48.2	3535
menopause polygenic score (count of menopause age increasing effect alleles, see table A1)	16.26	2.53	9	25	4698
menopause weighted polygenic score	3.86	.62	2.02	6.5	4698
endometriosis polygenic score (count of endometriosis increasing risk alleles, see table A2)	61.57	7.54	33	85	4698

Table 2. Do menopause/endometriosis SNPs predict childlessness?

Panel A: SNPs increasing menopause age. Logistic regression of childlessness at age 45.

	b (se)	b (se)	b (se)	b (se)	b (se)
birth year -1900	-3.635 (.68)***	-3.636 (.68)***	-3.568 (.77)***	-3.609 (.68)***	-3.258 (.79)***
squared (birth year -1900)	.354 (.07)***	.354 (.07)***	.354 (.07)***	.352 (.07)***	.353 (.07)***
education in years	.022 (.02)	.022 (.02)	.022 (.02)	-.226 (.14)~	-.272 (.14)~
polygenic score of *menopause* increasing effect alleles		.002 (.02)	.026 (.13)	-.204 (.11)~	-.120 (.15)
interaction with birth year			-.004 (.02)		-.022 (.02)
interaction with education				.015 (.01)~	.018 (.01)*
intercept	6.631 (1.66)***	6.597 (1.68)***	6.225 (2.63)*	9.872 (2.43)***	8.546 (2.86)**
N	4508	4508	4508	4508	4508

Panel B: SNPs increasing risk of endometriosis. Logistic regression of childlessness at age 45.

	b (se)	b (se)	b (se)	b (se)	b (se)
birth year -1900	-3.635 (.68)***	-3.659 (.68)***	-2.933 (.85)***	-3.670 (.68)***	-3.268 (.87)***
squared (birth year -1900)	.354 (.07)***	.356 (.07)***	.352 (.07)***	.357 (.07)***	.354 (.07)***
education in years	.022 (.02)	.023 (.02)	.022 (.02)	.415 (.17)*	.370 (.18)*
polygenic score of *endometriosis* risk alleles		-.011 (.01)~	.049 (.04)	.075 (.04)*	.098 (.05)*
interaction with birth year			-.011 (.01)		-.006 (.01)
interaction with education				-.006 (.00)*	-.006 (.00)~
intercept	6.631 (1.66)***	7.380 (1.71)***	3.585 (3.11)	2.138 (2.84)	.650 (3.46)
N	4508	4508	4508	4508	4508

~ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Table 3. Do menopause/endometriosis SNPs predict completed fertility?

Panel A: SNPs increasing menopause age. Poisson regression of number ever born at age 45.

	b (se)	b (se)	b (se)	b (se)	b (se)
birth year -1900	-3.635 (.68)***	.023 (.15)	-.034 (.17)	.023 (.15)	-.042 (.17)
squared (birth year -1900)	.354 (.07)***	-.008 (.02)	-.008 (.02)	-.008 (.02)	-.008 (.02)
education in years	.022 (.02)	-.002 (.00)	-.001 (.00)	-.001 (.03)	.006 (.03)
polygenic score of *menopause* increasing effect alleles		.000 (.00)	-.020 (.03)	.000 (.02)	-.016 (.03)
interaction with birth year			.004 (.01)		.004 (.01)
interaction with education				-.000 (.00)	-.000 (.00)
intercept	6.631 (1.66)***	.955 (.39)*	1.270 (.58)*	.954 (.53)~	1.212 (.62)~
N	4508	4508	4508	4508	4508

Panel B: SNPs increasing risk of endometriosis. Poisson regression of number ever born at age 45.

	b (se)	b (se)	b (se)	b (se)	b (se)
birth year -1900	-3.635 (.68)***	.024 (.15)	-.074 (.18)	.029 (.15)	-.007 (.19)
squared (birth year -1900)	.354 (.07)***	-.008 (.02)	-.008 (.02)	-.008 (.02)	-.008 (.02)
education in years	.022 (.02)	-.002 (.00)	-.001 (.00)	-.072 (.03)*	-.069 (.04)~
polygenic score of *endometriosis* risk alleles		.001 (.00)	-.008 (.01)	-.015 (.01)~	-.017 (.01)~
interaction with birth year			.002 (.00)		.001 (.00)
interaction with education				.001 (.00)*	.001 (.00)~
intercept	6.631 (1.66)***	.912 (.39)*	1.439 (.67)*	1.845 (.60)**	1.987 (.73)**
N	4508	4508	4508	4508	4508

Appendix I

Table A1. The 17 SNPs associated with age at menopause (Stolk, Perry et al. 2012; Perry, Corre et al. 2013) and descriptives of these SNPs in the LifeLines sample.

SNP id	chr.	reference allele	effect allele	effect allele frequency (mean)	homozygous for reference allele (%)	homozygous for effect allele (%)	heterozygous (%)	reported effect size per effect allele in (in years)(Stolk, Perry et al. 2012)
rs4246511	1	C	T	0.619	.478	.097	.425	.240
rs1635501	1	C	T	1.106	.199	.305	.495	.164
rs2303369	2	T	C	1.097	.196	.294	.510	.175
rs10183486	2	T	C	1.261	.136	.397	.467	.196
rs4693089	4	A	G	1.006	.247	.253	.499	.228
rs365132	5	G	T	0.935	.284	.219	.497	.287
rs2153157	6	G	A	1.055	.142	.198	.660	.165
rs1046089	6	A	G	1.275	.119	.394	.487	.213
rs2517388	8	T	G	0.379	.656	.036	.308	.262
rs12294104	11	C	T	0.298	.719	.017	.265	.225
rs2277339	12	G	T	1.945	.000	.945	.054	.380
rs4886238	13	G	A	0.585	.498	.084	.418	.170
rs2307449	15	G	T	1.286	.122	.408	.470	.184
rs10852344	16	T	C	0.835	.340	.175	.486	.168
rs11668344	19	G	A	1.258	.140	.398	.462	.416
rs12461110	19	A	G	1.301	.006	.308	.686	.158
rs16991615	20	G	A	0.051	.949	.000	.051	.948

Table A2. Top 100 endometriosis SNPs (Albertsen, Chettier et al. 2013). Continues on next page.

Chr	SNP	R=Replicated S=signal -=top 100 hit	Minor Allele	Odds Ratio per minor allele
1	rs4654783	S	A	1.21
1	rs3765351	-	G	1.14
1	rs2235529	S	A	1.29
1	rs2473241	-	A	1.17
1	rs882024	-	C	1.13
1	rs882025	-	A	1.13
1	rs2983118	-	G	1.14
1	rs1395455	-	A	1.12
1	rs2786485	-	A	1.16
1	rs4660584	-	A	1.15
1	rs1039871	-	G	1.14
2	rs1368087	-	G	1.13
2	rs12473304	-	A	1.18
2	rs4284854	-	A	1.17
2	rs1160581	-	G	1.16
2	rs1519754	S	C	1.20
2	rs6734792	S	G	1.20
2	rs1519761	S	G	1.20
2	rs6757804	S	G	1.20
2	rs1434094	-	A	1.12
2	rs6706330	-	A	0.90
2	rs6738749	-	A	1.09
2	rs10171524	-	A	0.89
3	rs11713777	-	G	1.24
3	rs2236951	r	G	1.18
3	rs4305418	-	G	1.15
3	rs907059	-	A	1.17
3	rs1510272	-	A	1.14
3	rs10513491	-	A	1.15

4	rs1373475	-	A	1.19
4	rs11724057	-	A	1.22
4	rs978335	-	A	1.30
4	rs3922934	-	A	1.17
4	rs17403181	-	A	0.90
4	rs6835945	r	A	0.86
4	rs17279486	-	G	0.86
5	rs12517129	-	A	1.13
5	rs11740761	-	A	1.14
5	rs12186488	-	A	1.14
5	rs4594818	-	G	1.20
5	rs2918439	-	A	0.87
6	rs2748359	-	A	1.13
6	rs426518	-	G	1.20
6	rs6916251	r	G	1.17
6	rs760794	r	A	1.17
6	rs2223361	r	A	1.17
6	rs2206034	r	A	1.17
6	rs6903595	r	A	1.18
6	rs6904518	r	G	1.19
6	rs6907340	S	A	1.20
6	rs11964747	-	A	1.13
6	rs3129304	-	G	1.13
6	rs3129303	-	G	1.13
6	rs711274	-	A	0.85
7	rs10265932	-	A	1.12
7	rs2270221	-	A	1.19
7	rs1860786	-	G	1.12
7	rs6462315	-	A	1.14
7	rs12701165	-	G	1.15
7	rs2429213	-	G	0.87
8	rs7816936	r	A	0.85
9	rs1330383	r	A	1.17
9	rs10975519	r	A	1.19
9	rs1332290	r	A	1.18
9	rs1048274	r	A	1.17
9	rs10815398	r	C	1.16
9	rs10815402	r	A	1.18
9	rs2492813	-	G	1.65
9	rs815845	-	C	1.13
9	rs10739696	-	C	0.91
9	rs4836579	-	G	0.91
9	rs10760500	-	G	0.90
10	rs1875005	-	A	0.88
10	rs2942366	r	A	1.15
10	rs10508881	S	A	1.18
10	rs11193561	-	A	1.18
10	rs17608302	-	A	0.89
11	rs4910169	-	G	1.17
11	rs7129273	-	G	1.15
11	rs7106873	-	A	0.87
11	rs10765405	-	G	0.90
12	rs12426819	-	G	0.88
12	rs7963889	-	G	1.11
12	rs10859856	r	G	1.16
12	rs3596	r	G	1.16
12	rs1362969	-	A	0.86
12	rs10431397	-	A	1.74
12	rs2138077	-	A	1.13
12	rs1533352	-	G	1.12
12	rs10847559	-	A	1.12
13	rs9579955	-	G	0.83
13	rs1512883	-	A	1.15
14	rs10132077	r	C	1.19
14	rs1268843	r	A	1.19
15	rs2445751	-	A	1.12
17	rs12449465	r	G	1.25
20	rs6139282	-	A	1.35
20	rs8050	-	A	0.68
21	rs909182	-	A	1.10
22	rs5767685	-	G	1.29

Note: SNP numbers rs882024, rs1519761, rs2236951, rs6916251, rs6904518, rs1048274, rs2492813, rs4910169, rs7129273, rs7106873, and rs10765405 are not in polygenic score as not in LifeLines genotypic data. SNPs associated with lower risk were flipped.

Appendix II: Polygenic menopause score & self-reported menopause & fertility

We first examined whether the age at menopause SNPs predict self-reported age at menopause in the LifeLines sample (see table A3 in the appendix). Note that LifeLines was part of the original GWAS on menopause(Stolk, Perry et al. 2012), so it would be strange not to find a relationship. We are well aware of potential problems using the same sample twice in risk prediction, future versions of this paper will deal with this issue. To our knowledge, endometriosis was not (yet) measured in LifeLines, so we cannot repeat this exercise for endometriosis. Second, we examine whether the reported age at menopause is related to fertility (see table 3). Additionally in this step, we also examine the age at last birth of a child. We suppose that -as a later menopausal age widens the reproductive window- reported menopause age should be positively related to fertility (increased number of children, decrease odds of childlessness, increased age at last birth).

Table A3 examines whether a polygenic score based on the 17 menopause SNPs predicts self-reported age at menopause. The table shows two versions of this score one based on a simple count of the effect alleles and the second weighting the alleles by effect size. It is clear from both the Cox survival model and the simple OLS regression model that the polygenic scores predict self-reported age at menopause. The weighted score is a much better predictor, but as LifeLines was part of the original GWAS, we decided not use this weighted score for the other analyses. Future versions of this paper will deal with this issue.

In table A4 we show that in post-menopausal women the self-reported age of menopause is related to their fertility. A later menopausal age is related to a slightly lower number of children, but not to childlessness. And a higher menopausal age is related to a later age of last child birth.

Table A3. Age at menopause SNPs predict reported age at menopause

	Cox Hazard-ratio's	OLS b	Cox Hazard-ratio's	OLS b
birth year -1900 /10	25.39***	15.17***	25.30***	15.19***
birth year -1900 /10 squared	0.76***	-1.80***	0.76***	-1.80***
menopause polygenic score	0.97***	0.12**		
menopause weighted polygenic score			0.76***	0.42**
constant		16.42***		16.42***
N	7687	2959	7687	2959
N failures	2959		2959	

legend: * p<0.05; ** p<0.01; *** p<0.001

Table A4. Reported age at menopause related to fertility

	number of children at 45 (OLS) b	childless at age 45 (logistic regression) b	age at last birth (OLS) b
birth year -1900 /10	-1.83***	2.63	-14.98***
birth year -1900 /10 squared	0.16***	-0.18	1.51***
educational level (years)	-0.01	0.08*	0.40***
reported age at menopause	0.01**	-0.01	0.07***
constant	7.13***	-11.88*	57.39***
N	2583	2583	1999

legend: * p<0.05; ** p<0.01; *** p<0.001

Acknowledgments

VIDI Grant 452-10-012, NWO/Dutch Science Foundation.

References

Albertsen, H. M., R. Chettier, et al. (2013). "Genome-Wide Association Study Link Novel Loci to Endometriosis." Plos One **8**(3): e58257.

Billari, F. C., H.-P. Kohler, et al. (2007). "Approaching the limit: Long-term trends in late and very late fertility." Population and Development Review **33**(1): 149-170.

Duncan, L. E. and M. C. Keller (2011). "A Critical Review of the First 10 Years of Candidate Gene-by-Environment Interaction Research in Psychiatry." Am J Psychiatry **168**(10): 1041-1049.

Hewitt, J. (2012). "Editorial Policy on Candidate Gene Association and Candidate Gene-by-Environment Interaction Studies of Complex Traits." Behavior Genetics **42**(1): 1-2.

Mirowsky, J. (2005). "Age at first birth, health, and mortality." Journal of health and social behavior **46**: 32-50.

Montgomery, G. W., K. T. Zondervan, et al. (2013). "The future for genetic studies in reproduction." Molecular Human Reproduction.

Nyholt, D. R., S.-K. Low, et al. (2012). "Genome-wide association meta-analysis identifies new endometriosis risk loci." Nat Genet **44**(12): 1355-1359.

Painter, J. N., C. A. Anderson, et al. (2011). "Genome-wide association study identifies a locus at 7p15.2 associated with endometriosis." Nat Genet **43**(1): 51-54.

Perry, J. R. B., T. Corre, et al. (2013). "A genome-wide association study of early menopause and the combined impact of identified variants." Human Molecular Genetics **22**(7): 1465-1472.

Stolk, L., J. R. B. Perry, et al. (2012). "Meta-analyses identify 13 loci associated with age at menopause and highlight DNA repair and immune pathways." Nat Genet **44**(3): 260-268.

Stolk, R. P., J. G. M. Rosmalen, et al. (2008). "Universal risk factors for multifactorial diseases." European Journal of Epidemiology **23**(1): 67-74.

Zhao, H. and Z.-J. Chen (2013). "Genetic association studies in female reproduction: from candidate-gene approaches to genome-wide mapping." Molecular Human Reproduction **19**(10): 644-654.