Supplementary material: A combined clinical and genetic model for predicting risk of ovarian cancer

Supplementary methods

Multiple imputation

We imputed nine complete datasets (von Hippel, 2020) using chained equations (linear regression for body mass index; logistic regression for ever had a full-term pregnancy, ever taken hormone replacement therapy and ever taken hormonal birth control; Poisson regression for number of full-term pregnancies; predictive mean matching (with 5 nearest neighbours) for duration of hormonal birth control use, age at menopause and duration of hormone replacement therapy use). Including the menopausal status variable in the multiple imputation introduced problems with perfect prediction of age at menopause; therefore, for the 22,670 (12.0%) unaffected women and 75 (10.6%) affected women with missing menopausal status, we considered women aged 50 years or less to be premenopausal and the rest to be postmenopausal before running the imputation.

Polygenic risk score

To calculate each woman's population-standardised PRS (Mealiffe et al., 2010, Conran et al., 2016), we first calculated the unscaled population average risk (μ) for each SNP as $\mu = (1-p)^2 + 2p(1-p)OR + p^2OR^2$, where p is the risk allele frequency and OR is the odds ratio per effect allele. Next, for each of the 36 SNPs, each woman's adjusted risk was calculated as $\frac{OR^N}{\mu}$, where N was the woman's number of effect alleles (missing SNPs were given an adjusted risk of 1). This ensured that each SNP's adjusted risk had a population mean equal to 1. Finally, the 36-SNP PRS (*snprisk*, as a relative risk) for each woman was the product of their adjusted risks for each of the SNPs.

Clinical risk score

The clinical risk score from Li et al. (2015) (as a relative risk) was calculated as $clinrisk = e^{(0.019 \times menop + 0.034 \times menopage + 0.086 \times hrt + 0.057 \times hrttime - 0.181 \times hbc - 0.034 \times hbctime - 0.308 \times birth - 0.094 \times birthn + 0.021 \times bmi)}$, where *menop*, *hrt*, *hbc* and *birth* were indicator variables for being menopausal, ever taking hormone replacement therapy, ever taking hormonal birth control and ever having had a full-term pregnancy, respectively. The *menopage* variable was age in years at menopause minus 50 for menopausal women and 0 for premenopausal women; *hrttime* was years of hormone replacement therapy use minus 2.5 for women who had taken hormone replacement therapy and 0 for women who had not; *hbctime* was years of hormonal birth control use minus 5 for women who had taken hormonal birth control and 0 for women who had not; *birthn* was number of full-term pregnancies minus 2 for parous women and 0 for non-parous women; *bmi* was body mass index (kg/m^2). We did not include the clinical model's term for having had one ovary removed because, in the UK Biobank, some of these women were unable to be distinguished from women with both ovaries removed; therefore, all women with a history of ovary removal were excluded from our analysis dataset. As a final step, we divided the clinical risk score by its mean to ensure that the mean clinical risk was equal to 1, that is, $clinrisk\_c = \frac{clinrisk}{1.236909}$.

Combined risk score

To create the combined risk score, we multiplied the PRS by the centred clinical risk score: *combrisk* = *snprisk* × *clinrisk_c*.

Absolute risks

For the calculation of 10-year risk for each woman (aged *b* years at her baseline assessment date), we first determined her population incidence from birth to age *b* years (*popincid*) and to age *b* + 10 years (*popincid10*). For each risk score (e.g. *combrisk*), we then

calculated each woman's cumulative risks to age *b* years and age *b* + 10 years as: $cumul = 1 - e^{-combrisk \times popincid}$ and $cumul10 = 1 - e^{-combrisk \times popincid10}$, respectively. For each woman, her expected survival in the next 10 years was $surv10 = e^{-mort10}$, where *mort10* was her expected mortality in the next 10 years. The 10-year absolute risk of ovarian cancer was then calculated as $absrisk10yr = \frac{(cumul10 - cumul) \times surv10}{(1 - cumul)}$.

For the recalibrated risk score (*newrisk*), we also calculated the full-lifetime risk for each woman as $fullliferisk = 1 - e^{newrisk \times popincidlife}$, where *popincidlife* was her cumulative incidence of ovarian cancer from birth to age 85 years.

Supplementary tables

Supplementary Table 1. Eligibility criteria for the current study

N eligible	Criteria	N dropped
502,413	Active UK Biobank participants (at 22 February 2022)	
273,185	Female and reported sex same as genetic sex	228,850
271,630	Aged 40–69 years at baseline assessment date	1555
219,894	Caucasian	51,736
219,039	No ovarian cancer at baseline assessment date	855
198,592	Have both ovaries at baseline assessment	20,447
198,256	Genotyping data available	336
198,247	Alive after six weeks of follow-up	9
198,244	No ovarian cancer after six weeks of follow-up	3
189,171	Unrelated individuals (≥3rd degree relatedness)	9073

Supplementary Table 2. Number of single-nucleotide polymorphisms genotyped for unaffected and affected women

Number of SNPs genotyped	Unaffected		Affected	
	N	%	N	%
32	4	0.0	0	0.0
33	229	0.1	0	0.0
34	3479	1.9	21	3.0
35	33,002	17.5	123	17.3
36	151,746	80.5	567	79.8

Note: SNP, single-nucleotide polymorphism

Supplementary Table 3. Baseline risk factors used in the clinical risk score for women who were unaffected and affected with ovarian cancer

Categorical risk factors	Unaffected		Affected	
	N	%	N	%
Number of full-term pregnancies				
0	34,356	18.2	141	19.8
1	24,658	13.1	103	14.5
2	85,423	45.3	309	43.5
3	33,286	17.7	124	17.4
4	8209	4.4	27	3.8
5+	2418	1.3	7	1.0
Missing	110	0.1	0	0.0
Hormonal birth control, ever used				
No	37,979	20.2	209	29.4
Yes	150,164	79.7	499	70.2
Missing	317	0.2	3	0.4
Menopause				
No	45,653	24.2	105	14.8
Yes	120,137	63.8	531	74.7
Missing	22,670	12.0	75	10.6
Hormone replacement therapy, ever used				
No	128,016	67.9	430	60.5
Yes	59,990	31.8	281	39.5
Missing	454	0.2	0	0.0
Continuous risk factors	**Unaffected**		**Affected**	
	Mean	**SD**	**Mean**	**SD**
Body mass index (kg/m^2)	26.9	5.1	27.4	5.1
Duration of hormonal birth control use (years)*	10.6	7.5	8.7	6.9
Age at menopause (years)*	49.9	4.9	50.2	5.0
Duration of hormone replacement therapy use (years)*	7.4	5.2	8.1	5.3

Note: SD, standard deviation; * in women who answered yes to the corresponding indicator question.

Supplementary Table 4. Refined estimates for the final model using the PRS and clinical risk score in the full dataset

Relative risk score	ln HR	95% CI	*P* value
ln PRS (*snprisk*)	0.898	0.673, 1.124	<0.001
ln clinical risk score (*clinrisk_c*)	0.454	0.260, 0.647	<0.001

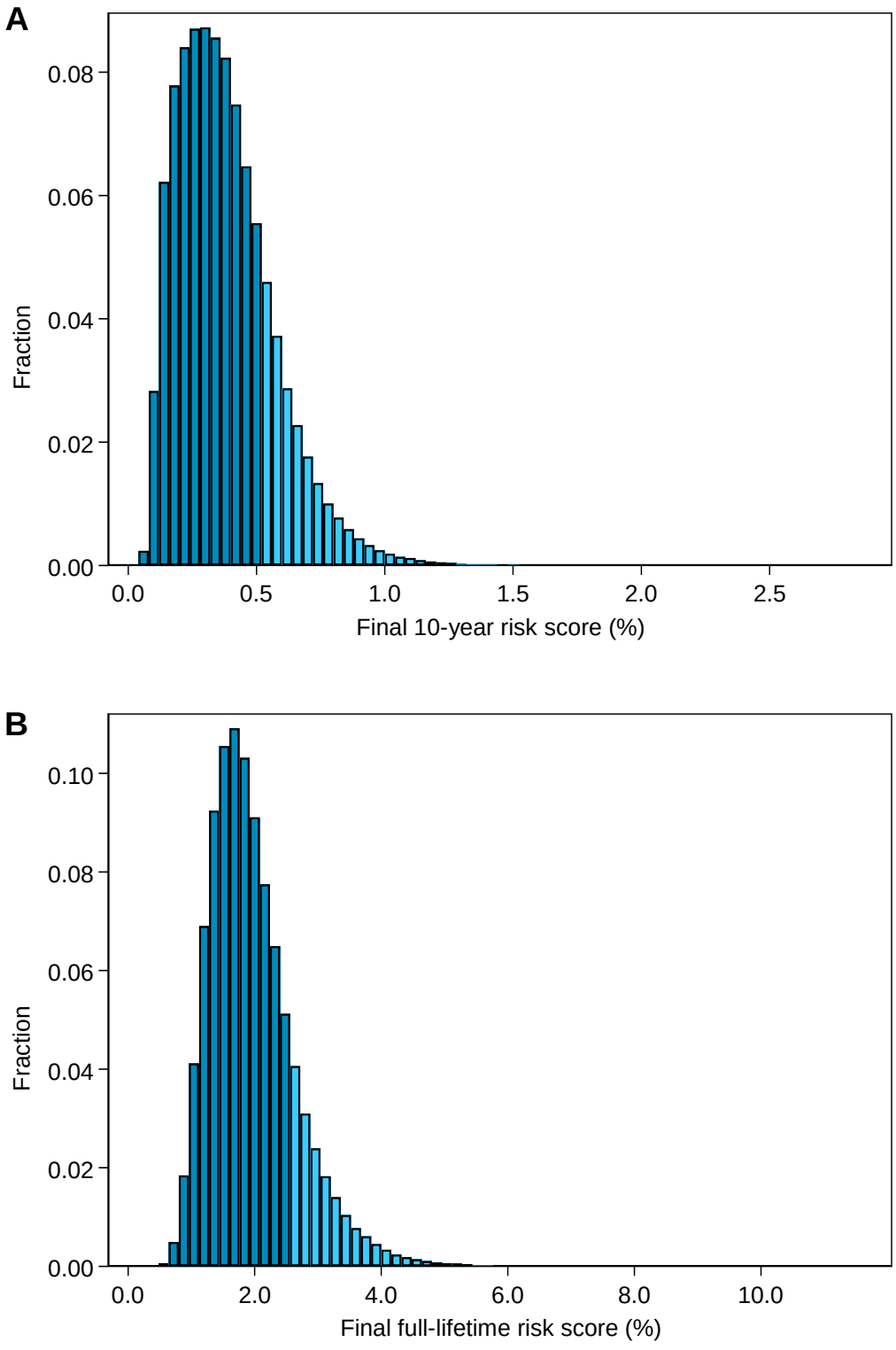
Note: CI, confidence interval; HR, hazard ratio; ln, natural logarithm; PRS, polygenic risk score; SNP, single-nucleotide polymorphism.

For the final model, the relative risk is calculated as follows:

$$finalrisk = e^{0.898 \times \ln(snprisk) + 0.454 \times \ln(clinrisk\_c)},$$

where *snprisk* and *clinrisk_c* are calculated as described in the Supplementary Methods. This can then be used with appropriate population incidence rates to calculate the 10-year (or *n*-year) risk or full-lifetime risk of ovarian cancer, as described in the Methods.

Supplementary figure



Supplementary Figure 1. Distribution of the final (A) 10-year risk score and (B) full-lifetime risk score in the full dataset. The light-blue sections correspond to the top quintile of risk. For the 10-year risk, the median is 0.3%, and for the full-lifetime risk, the median is 1.8%.

Supplementary references

Conran CA, Na R, Chen H, Jiang D, Lin X, Zheng SL et al. (2016). Population-standardized genetic risk score: the SNP-based method of choice for inherited risk assessment of prostate cancer. *Asian J Androl* **18**(4):520-524. doi: 10.4103/1008-682X.179527.

Li K, Husing A, Fortner RT, Tjonneland A, Hansen L, Dossus L et al. (2015). An epidemiologic risk prediction model for ovarian cancer in Europe: the EPIC study. *Br J Cancer* **112**(7):1257-1265. doi: 10.1038/bjc.2015.22.

Mealiffe ME, Stokowski RP, Rhees BK, Prentice RL, Pettinger M, Hinds DA (2010). Assessment of clinical validity of a breast cancer risk model combining genetic and clinical information. *J Natl Cancer Inst* **102**(21):1618-1627. doi: 10.1093/jnci/djq388.

von Hippel PT (2020). How many imputations do you need? A two-stage calculation using a quadratic rule. *Sociological Methods and Research* **49**(3):699-718. doi: 10.1177/0049124117747303.