OPEN

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

Gillian S. Dite[a], Erika Spaeth[b], Nicholas M. Murphy[a] and Richard Allman[a]

Objective **Women with a family history of ovarian cancer or a pathogenic or likely pathogenic gene variant are at high risk of the disease, but very few women have these risk factors. We assessed whether a combined polygenic and clinical risk score could predict risk of ovarian cancer in population-based women who would otherwise be considered as being at average risk.**

Methods **We used the UK Biobank to conduct a prospective cohort study assessing the performance of 10-year ovarian cancer risks based on a polygenic risk score, a clinical risk score and a combined risk score. We used Cox regression to assess association, Harrell's C-index to assess discrimination and Poisson regression to assess calibration.**

Results **The combined risk model performed best and problems with calibration were overcome by recalibrating the model, which then had a hazard ratio per quintile of risk of 1.338 [95% confidence interval (CI), 1.152–1.553], a Harrell's C-index of 0.663 (95% CI, 0.629–0.698) and overall calibration of 1.000 (95% CI, 0.874–1.145). In the refined model with estimates based on the entire dataset, women in the top quintile of 10-year risk were at 1.387 (95% CI, 1.086–1.688) times increased risk, while women in the top quintile of full-lifetime risk were at 1.527 (95% CI, 1.187–1.866) times increased risk compared with the population.**

Conclusion **Identification of women who are at high risk of ovarian cancer can allow healthcare providers and patients to engage in joint decision-making discussions around the risks and benefits of screening options or risk-reducing surgery.** ***European Journal of Cancer Prevention*** **32: 57–64 Copyright © 2022 The Author(s). Published by Wolters Kluwer Health, Inc.**

European Journal of Cancer Prevention 2023, **32**:57–64

Keywords: absolute risk, clinical risk, ovarian cancer, polygenic risk, risk prediction

[a]Genetic Technologies Limited, Fitzroy, Victoria, Australia and [b]Phenogen Sciences Inc, Charlotte, North Carolina, USA

Correspondence to Gillian S. Dite, PhD, Genetic Technologies Limited, 60–66 Hanover St, Fitzroy VIC 3065, Australia
Tel: +61 3 8412 7000; e-mail: gillian.dite@gtglabs.com

Received 18 August 2022 **Accepted** 6 September 2022

Introduction

Despite ovarian cancer being one of the less common cancers in women, it is one of the top five cancer-associated causes of death in women (American Cancer Society, 2022). Ovarian cancer presents with few and often non-specific symptoms that impede early detection efforts. In Western countries, high-grade serous carcinomas account for two-thirds of ovarian cancers and are clinically aggressive neoplasms that are usually diagnosed at an advanced stage (Koshiyama *et al.*, 2017). Women are currently considered high risk based on their family history of ovarian cancer or if they have a pathogenic or likely pathogenic variant in *BRCA1*, *BRCA2* or another high-penetrance variant associated with ovarian cancer (National Comprehensive Cancer Network, 2022).

The United States' National Comprehensive Cancer Network (2022) guidelines use a 10% full-lifetime risk threshold for a recommendation to discuss risk-reducing salpingo-oophorectomy. However, for women who are carriers of moderate-penetrance variants (e.g. *PALB2*) and have a family history of ovarian cancer, this surgery may be considered at 5% full-lifetime risk, or almost four times the average risk of developing ovarian cancer. For high-risk women who choose not to undergo surgery, transvaginal ultrasound and screening for cancer antigen 125 (with the Risk of Ovarian Cancer Algorithm [Skates, 2012] to interpret results) are available.

Traditionally, women have been defined as high risk based on their family history of ovarian cancer or if they have a pathogenic or likely pathogenic variant in *BRCA1*, *BRCA2* or another high-penetrance variant that is associated with an increased risk of developing ovarian cancer. In these high-risk women, screening improved early detection (Rosenthal *et al.*, 2017; Skates *et al.*, 2017), while in the general population, screening identified more early-stage cancers and fewer late-stage cancers but did not reduce mortality (Buys *et al.*, 2011; Menon *et al.*, 2021).

In the general population, a family history of ovarian cancer (La Vecchia, 2017; Hu *et al.*, 2021) and having a pathogenic variant (Hu *et al.*, 2021) are rare. For the general population, established clinical risk scores can be used

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (www.eurjcancerprev.com).

This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

0959-8278 Copyright © 2022 The Author(s). Published by Wolters Kluwer Health, Inc. DOI: 10.1097/CEJ.0000000000000771

for risk stratification to better target screening to women at increased risk, and it is possible that the risk stratification could be improved by including genetic information. Genome-wide association studies have identified common genetic variants associated with risk of epithelial ovarian cancer (Song *et al.*, 2009; Phelan *et al.*, 2017) that have been used to construct polygenic risk scores (PRSs; Yang *et al.*, 2018; Dareng *et al.*, 2022). While many of these PRSs have focused on modifying pathogenic or likely pathogenic variant carrier risk, the low prevalence of carriers in the population limits the benefits to a small number of women (Kuchenbaecker *et al.*, 2017; Barnes *et al.*, 2020). However, Dareng *et al.* (2022) developed several population-based PRSs with varying numbers of single-nucleotide polymorphisms (SNPs) using a range of statistical methods.

Unlike risk models that have combined the effect of pathogenic or likely pathogenic variants with polygenic risk and clinical risk factors (Lee *et al.*, 2021), here we focus on risk prediction for women in the general population. In this article, we use the population-based UK Biobank (Sudlow *et al.*, 2015; Bycroft *et al.*, 2018) to assess the utility of combining an established clinical model developed by Li *et al.* (2015) with a PRS based on a panel of 36 SNPs developed by Dareng *et al.* (2022) to predict risk of ovarian cancer in the general population.

Methods

UK Biobank

From 2006–2009, the UK Biobank recruited over 500 000 participants from England, Wales and Scotland (Sudlow *et al.*, 2015; Bycroft *et al.*, 2018). The baseline assessment comprised a touch-screen questionnaire, a face-to-face interview and physical measurements to collect phenotypic data. Genomic and biomarker data have been derived from the biological samples that were collected at the baseline assessment. Outcome data are available through linkage to cancer registries, death registries and hospital data for the whole cohort.

Ovarian cancer was identified based on self-reported data and linked cancer registry data. Ovarian cancers were incident if diagnosed after the baseline assessment date and prevalent otherwise. Linked hospital data were used to identify women who had one or both ovaries removed after the baseline assessment date and linked death registry data were used to identify women who died during the follow-up period.

Eligibility

Eligible participants were female, genetically Caucasian, aged 40–69 years at their baseline assessment date and had not withdrawn their participation as of 22 February 2022. Women who had ovarian cancer diagnosed before their baseline assessment date, had one or both ovaries removed before their baseline assessment date, had no SNP data available, or had died or been diagnosed with ovarian cancer within the first 6 weeks of follow-up were excluded from all analyses. To limit the final dataset to unrelated women (i.e. not closer than third-degree relatedness), we used the ukb_gen_samples_to_remove function of the R package ukbtools (Hanscombe *et al.*, 2019). Of the 189 171 women in the final analysis dataset, 711 (0.4%) were diagnosed with incident ovarian cancer during the 10-year follow-up period. Details of the eligibility criteria are shown in Supplementary Table 1, Supplemental Digital Content 1, *http://links.lww.com/EJCP/A368*.

Risk scores

We extracted genotypes for Dareng *et al.*'s (2022) panel of 36 SNPs from the UK Biobank's SNP imputation dataset using Plink version 1.9. (Chang *et al.*, 2015, Purcell and Chang, 2020). The population-standardised PRS (as a relative risk; Mealiffe *et al.*, 2010; Conran *et al.*, 2016) was constructed using Dareng *et al.*'s (2022) European population estimates of the odds ratio per effect allele and effect allele frequency.

Some of the risk factors required for the clinical risk score had missing data and a complete-case analysis would have resulted in the loss of 47 307 (25.1%) unaffected women and 181 (25.5%) affected women. We, therefore, used multiple imputation to ensure that all eligible women were included in the analyses. For the clinical risk score (as a relative risk), we applied the selected model developed by Li *et al.* (2015) to data from the baseline assessment. We centred the clinical risk score by dividing it by its mean (1.237). To create the combined risk score, we multiplied the PRS by the centred clinical risk score.

We calculated absolute 10-year risk for each of the risk scores using age-specific (in 5-year groups) and calendar year-specific ovarian cancer incidence rates for England (Office for National Statistics, 2019) as population reference rates. For the 10-year risk, we also used a competing mortality adjustment using age-specific (in 5-year groups) and calendar year-specific mortality rates for causes of death other than ovarian cancer (Office of National Statistics, 2016, 2019). For the recalibrated risk score, we also calculated full-lifetime risk. Full details of the calculation of the risk scores and the multiple imputation are in the Supplementary Methods, Supplemental Digital Content 1, *http://links.lww.com/EJCP/A368*.

Statistical analysis

Each woman's follow-up began at the date of her baseline assessment and stopped at whichever came first of date of completing 10 years of follow-up, date of diagnosis of ovarian cancer, date of ovary removal, date of death or 31 July 2019 (to which linkage to cancer registries is complete). Using the multiple imputation datasets, we used Cox regression with age as the time axis

to estimate the hazard ratio (HR) per quintile of risk for the clinical risk score, the PRS and the combined risk score. Discrimination was assessed using Harrell's C-index. Calibration of the risk scores was assessed using Poisson regression to obtain standardised incidence ratios (SIRs; by multiplying the incidence rate ratios by the constant term) of the observed number of ovarian cancer cases divided by the number expected by the 10-year risk scores, overall and by quintile of risk. We also used Poisson regression to obtain SIRs of the number of cancers seen in the UK Biobank divided by the number expected using population incidence rates overall, by 10-year age group and for each quintile of risk.

To recalibrate the combined risk model, we randomly divided the imputed data into 70% training and 30% testing datasets that were balanced for affected status. In the 70% training dataset, we used Cox regression to obtain HRs for the natural logarithm of the PRS and the natural logarithm of the clinical risk score. We used the natural logarithm of the HRs to predict a new risk score for all women and assessed the performance of the new risk score in the 30% testing dataset using the methods described above for the individual risk scores.

As a final step, we refined the estimates for the recalibrated model using all available data and used these to calculate the 10-year and full-lifetime risks for all participants. We then used Poisson regression to obtain SIRs of the number of cancers seen in the UK Biobank and the number expected using population incidence rates for each quintile of 10-year risk and each quintile of full-lifetime risk. We used Stata version 17.0 (StataCorp, 2021) for the analyses; all statistical tests were two-sided and P values <0.05 were considered nominally statistically significant.

Ethics approval

The UK Biobank has Research Tissue Bank approval (REC #11/NW/0382) that covers analysis of data by approved researchers. All participants provided written informed consent to the UK Biobank before data collection began. This research has been conducted using the UK Biobank resource under Application Number 47401.

Data availability

The data for this study were provided by the UK Biobank and we do not have permission to share the data. Researchers wishing to access the data used in this study can apply directly to the UK Biobank at https://www.ukbiobank.ac.uk/register-apply. Stata 17.0 code for the data management and analysis is available from the corresponding author for non-commercial purposes.

Results

In the final dataset, 188 460 women were unaffected and 711 had incident ovarian cancer during the 10-year follow-up period. Unaffected women had a mean age of 56.9 (standard deviation [SD] = 7.9) years at their baseline assessment, while affected women had a mean age of 59.9 (SD = 7.2) years at their baseline assessment and a mean age of 65.0 (SD = 7.8) years at their diagnosis of ovarian cancer. Unaffected women had a mean of 9.7 (SD = 0.9) years of follow-up until their censor date, while affected women had a mean of 6.7 (SD = 3.4) years of follow-up until their diagnosis of ovarian cancer.

Overall, 152 313 (80.5%) women had all 36 SNPs genotyped and 33 125 (17.5%) were missing one SNP; see Supplementary Table 2, Supplemental Digital Content 1, *http://links.lww.com/EJCP/A368*. The mean PRS was 1.01 (SD = 0.34) for unaffected women and 1.11 (SD = 0.37) for affected women. The mean of the centred clinical risk score (before multiple imputation) was 1.00 (SD = 0.43) in unaffected women and 1.16 (SD = 0.45) in affected women. The characteristics of the unaffected and affected women for the risk factors in the clinical risk score are shown in Supplementary Table 3, Supplemental Digital Content 1, *http://links.lww.com/EJCP/A368*.

The number of ovarian cancers seen in 10 years of follow-up was close to the number expected using population rates (SIR = 0.944; 95% CI, 0.877–1.016; $P = 0.1$, where P is for the null hypothesis that the SIR = 1). This was also evident when stratified by 10-year age group: for women aged 40–49 years, SIR = 0.922 (95% CI, 0.731–1.112; $P = 0.4$); for women aged 50–59 years, SIR = 0.882 (95% CI, 0.762–1.002; $P = 0.07$); for women aged 60–69 years, SIR = 0.984 (95% CI, 0.889–1.079; $P = 0.7$).

Risk scores

Table 1 shows that the 10-year risk scores were strongly associated with risk of ovarian cancer and that Harrell's C-index of the combined risk score was slightly higher than those for the PRS and the clinical risk score. Table 1 also shows that the PRS was well calibrated, overall and by quintile of risk, with no trend evident ($P = 0.6$). The clinical and combined risk scores were both well calibrated overall, but examination of the SIRs by quintile of risk shows that both had a problem with risk dispersion (P for trend 0.004 and 0.01, respectively), with overestimation of risk in the highest quintile.

Recalibration

We created a recalibrated combined model using the 70% training dataset. For the natural logarithm of the PRS, the HR was 2.777 (95% CI, 1.741–2.976; $P < 0.001$), and for the natural logarithm of the clinical relative risk, the HR was 1.467 (95% CI, 1.169–1.841; $P = 0.001$). Table 2 shows the performance of the new 10-year risk score in the 30% testing dataset. The new model had improved performance in terms of the HR per quintile of risk and Harrell's C-index, and the problem with

Table 1 Performance of the polygenic risk score, clinical and combined 10-year risks in the full dataset

Association	HR per quintile of risk	95% CI	*P* value
PRS 10-year risk	1.156	1.081–1.236	<0.001
Clinical 10-year risk	1.092	1.004–1.188	0.04
Combined 10-year risk	1.268	1.178–1.365	<0.001
Discrimination	**Harrell's C-index**	**95% CI**	
PRS 10-year risk	0.623	0.603–0.642	
Clinical 10-year risk	0.622	0.603–0.641	
Combined 10-year risk	0.647	0.628–0.666	
Calibration	**SIR**	**95% CI**	***P* value***
PRS 10-year risk, overall	1.004	0.933–1.080	0.9
Quintile 1 (median=0.19%)	0.877	0.653–1.101	0.3
Quintile 2 (median=0.28%)	1.057	0.860–1.255	0.6
Quintile 3 (median=0.36%)	1.047	0.874–1.220	0.6
Quintile 4 (median=0.45%)	0.923	0.777–1.068	0.3
Quintile 5 (median=0.61%)	1.049	0.918–1.179	0.5
Clinical 10-year risk, overall	0.970	0.901–1.044	0.4
Quintile 1 (median=0.14%)	1.183	0.854–1.512	0.2
Quintile 2 (median=0.23%)	1.118	0.862–1.374	0.3
Quintile 3 (median=0.34%)	0.958	0.766–1.149	0.7
Quintile 4 (median=0.46%)	1.124	0.958–1.291	0.1
Quintile 5 (median=0.68%)	0.818	0.711–0.925	0.003
Combined 10-year risk, overall	0.960	0.892–1.033	0.3
Quintile 1 (median=0.12%)	1.260	0.908–1.610	0.1
Quintile 2 (median=0.21%)	0.936	0.694–1.178	0.6
Quintile 3 (median=0.31%)	1.066	0.855–1.277	0.5
Quintile 4 (median=0.45%)	1.057	0.899–1.214	0.5
Quintile 5 (median=0.73%)	0.843	0.744–0.942	0.005

CI, confidence interval; HR, hazard ratio; PRS, polygenic risk score; SIR, standardised incidence ratio; SNP, single-nucleotide polymorphism. The SIR is the observed number of ovarian cancer cases divided by the number expected by the 10-year risk prediction scores.
**P* value for test that SIR=1.

Table 2 Performance of the new 10-year risk score in the 30% testing dataset

Association	HR per quintile of risk	95% CI	*P* value
New 10-year risk score	1.338	1.152–1.553	<0.001
Discrimination	**Harrell's C-index**	**95% CI**	
New 10-year risk score	0.663	0.629–0.698	
Calibration	**SIR**	**95% CI**	***P* value***
Overall	1.000	0.874–1.145	1.0
Quintile 1 (median=0.16%)	0.592	0.203–0.982	0.1
Quintile 2 (median=0.26%)	0.843	0.468–1.218	0.4
Quintile 3 (median=0.35%)	0.920	0.589–1.250	0.6
Quintile 4 (median=0.44%)	1.075	0.759–1.391	0.6
Quintile 5 (median=0.60%)	1.133	0.888–1.379	0.3

CI, confidence interval; HR, hazard ratio; SIR, standardised incidence ratio. The SIR is the observed number of ovarian cancer cases divided by the number expected by the 10-year risk prediction score.
**P* value for test that SIR=1.

overestimation of risk in the highest quintile of risk was overcome. There was, however, evidence for a trend of overestimation of risk in the lower quintile to underestimation of risk in the highest quintile (P for trend = 0.03). For the new risk score expressed as a full-lifetime risk, the HR per quintile of risk was 1.301 (95% CI, 1.170–1.446; $P < 0.001$) and Harrell's C-index was 0.634 (95% CI, 0.599–0.670).

Table 3 shows the SIRs and 95% confidence intervals for quintiles of the new 10-year risk score compared with population incidence rates in the 30% testing dataset. There was clear evidence that the risk of ovarian cancer increased from the lowest to the highest quintiles of the new 10-year risk score (P for trend <0.001). Table 3 also shows similar results for the new full-lifetime risk score, with a clear trend of increased risk from lowest to highest quintile (P for trend <0.001).

Final model

After refining estimates using all available data, for the natural logarithm of the PRS, the HR was 2.456 (95% CI, 1.960–3.077; $P < 0.001$), and for the natural logarithm of the clinical relative risk, the HR was 1.574 (95% CI, 1.297–1.909; $P < 0.001$). Details of the calculation of the final relative risk score are shown in Supplementary Table 4, Supplemental Digital Content 1, *http://links.lww.com/EJCP/A368*. Table 4 and Fig. 1 show the SIRs

Table 3 Standardised incidence ratios for quintiles of the new 10-year and full-lifetime risk prediction scores compared with population incidence rates in the 30% testing dataset

Risk score	SIR	95% CI	*P* value*
New 10-year risk			
Quintile 1 (median=0.16%)	0.403	0.138–0.667	0.007
Quintile 2 (median=0.26%)	0.640	0.355–0.925	0.05
Quintile 3 (median=0.35%)	0.741	0.475–1.007	0.1
Quintile 4 (median=0.44%)	0.987	0.697–1.277	0.9
Quintile 5 (median=0.60%)	1.387	1.086–1.688	0.003
New full-lifetime risk			
Quintile 1 (median=1.27%)	0.345	0.155–0.536	<0.001
Quintile 2 (median=1.58%)	0.769	0.494–1.043	0.1
Quintile 3 (median=1.85%)	0.901	0.603–1.198	0.5
Quintile 4 (median=2.17%)	0.954	0.663–1.245	0.8
Quintile 5 (median=2.75%)	1.504	1.166–1.843	<0.001

CI, confidence interval; SIR, standardised incidence ratio. The SIR is the observed number of ovarian cancer cases divided by the number expected by the population incidence rates.
**P* value for test that SIR=1.

Table 4 Standardised incidence ratios for quintiles of the final 10-year and full-lifetime risk prediction scores compared with population incidence rates in the full dataset

Risk score	SIR	95% CI	*P* value*
Final 10-year risk			
Quintile 1 (median=0.16%)	0.439	0.167–0.712	0.009
Quintile 2 (median = 0.25%)	0.636	0.351–0.920	0.05
Quintile 3 (median=0.34%)	0.670	0.438–0.962	0.06
Quintile 4 (median=0.44%)	1.007	0.716–1.299	1.0
Quintile 5 (median=0.62%)	1.395	1.094–1.697	0.003
Final full-lifetime risk			
Quintile 1 (median=1.20%)	0.337	0.142–0.532	<0.001
Quintile 2 (median=1.54%)	0.745	0.464–1.026	0.1
Quintile 3 (median=1.83%)	0.862	0.567–1.157	0.4
Quintile 4 (median=2.19%)	0.983	0.688–1.277	0.9
Quintile 5 (median=2.84%)	1.527	1.187–1.866	<0.001

CI, confidence interval; SIR, standardised incidence ratio. The SIR is the observed number of ovarian cancer cases divided by the number expected by the population incidence rates.
**P* value for test that SIR=1.

and 95% CIs for quintiles of the final 10-year risk score and final full-lifetime risk score compared with population incidence rates. For both risk scores, there is a clear increase in risk from lowest to highest quintile (both *P* for trend <0.001). Distributions of the final 10-year risk and final full-lifetime risk scores are shown in Supplementary Fig. 1, Supplemental Digital Content 1, *http://links.lww.com/EJCP/A368*.

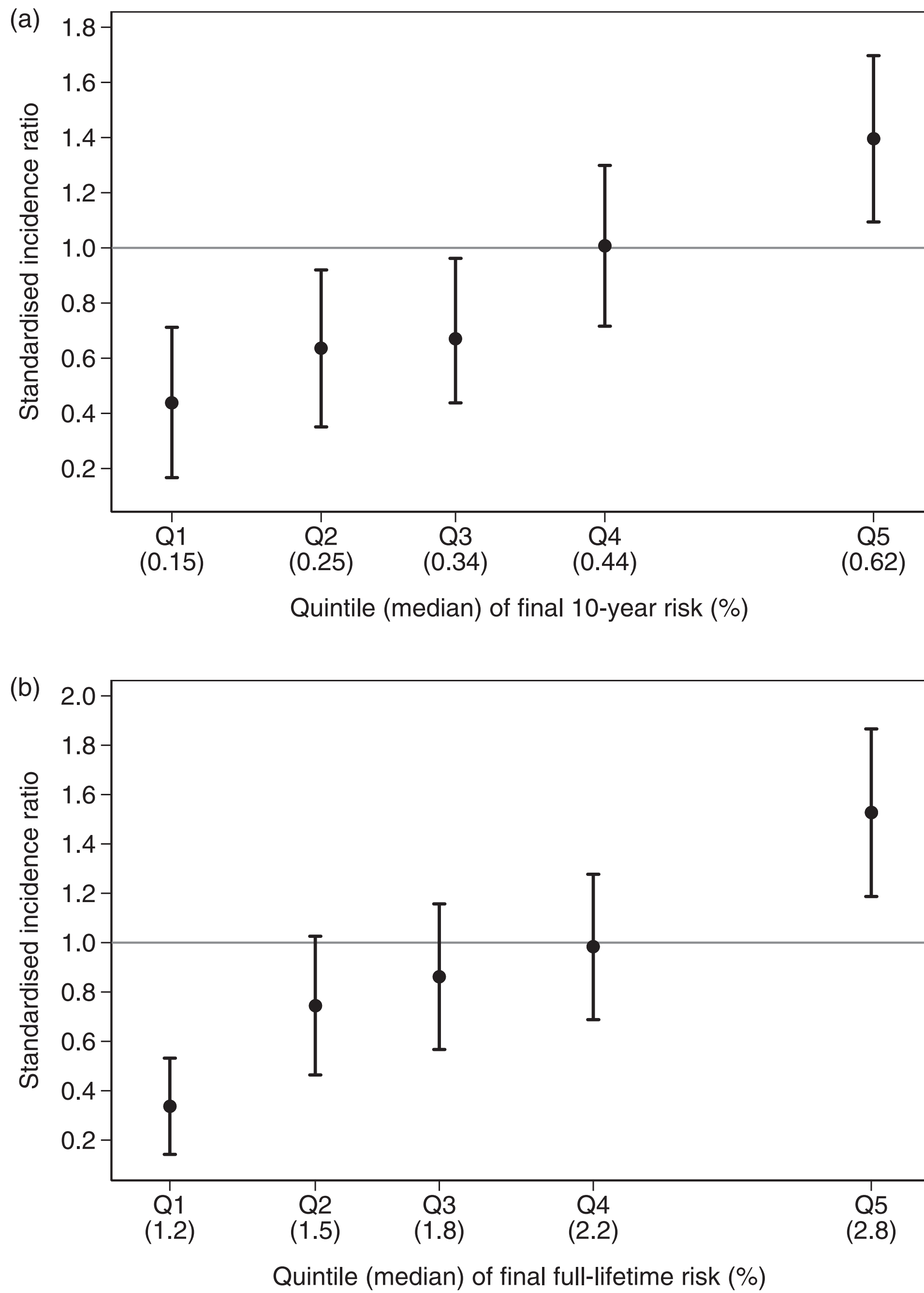
Discussion

A combined risk score incorporating a PRS and a clinical risk score has improved discrimination of ovarian cancer compared with the component risk scores. Recalibration addressed the problem with dispersion of risk and the overestimation of risk in the highest quintile of risk. The final (recalibrated) risk score was able to stratify women who would normally be considered to be at average risk such that women in the highest quintile of risk were at around 40% and 50% increased risk for the final 10-year risk score and final full-lifetime risk score, respectively.

In the UK Biobank, 2.48% of women had a full-lifetime risk of ≥5% and 2.22% had a full-lifetime risk of ≥10%; these are the thresholds used to initiate a discussion of risk-reducing salpingo-oophorectomy in carriers of moderate-penetrance variants and the general population, respectively. (Carriers of high-penetrance variants automatically qualify.) Using the recalibrated combined risk score, we can identify 20% of women who are at substantially increased risk of ovarian cancer using either the 10-year risk score (risks range from 0.51% to 2.89%) or the full-lifetime risk score (risks range from 2.42% to 11.69%); see Supplementary Fig. 1, Supplemental Digital Content 1, *http://links.lww.com/EJCP/A368*. For full-lifetime risk, this 20% of women – who would normally be considered at average risk – are at 1.9–9.1 times (median 2.2 times) the full-lifetime risk of the general population and may be ideal candidates for targeted screening.

Ovarian cancer is often detected at a late stage that is associated with more aggressive disease and higher cancer-associated mortality rates (Menon *et al*., 2021). Current screening options could be considered for high-risk women, despite the lack of evidence that mortality is reduced (Rosenthal *et al*., 2017; Skates *et al*., 2017). There is clear evidence that screening increases detection rates of early-stage cancers versus late-stage cancers (Rosenthal *et al*., 2017). Furthermore, many

Fig. 1



Standardised incidence ratios and 95% confidence intervals for the observed number of cancers compared with population incidence rates in the full dataset for quintiles of the (a) final 10-year risk and (b) final full-lifetime risk.

advances have been made in ovarian cancer treatment in the past 10 years (McMullen *et al.*, 2021), and mortality rates could indeed be reduced if early-stage cancer were identified. Finally, novel screening modalities, including testing for circulating tumour DNA are now commercially available options (Liu *et al.*, 2020), for which the ability to risk stratify the general population will likely improve the positive predictive value of those assays.

In this study we excluded prevalent ovarian cancer from all analyses, thereby avoiding the potential for incidence–prevalence bias to affect the performance of the risk scores. This also ensured that, for the clinical risk

factors, the temporal relationship between exposure and outcome was maintained. We have highlighted the importance of using a well-calibrated model for ovarian cancer risk prediction. With the recalibration, we eliminated the overestimation of risk in women at the higher end of the risk spectrum.

Limitations to the study include the lack of ethnic diversity available in the UK Biobank. Nevertheless, for now, the model can be used with ethnicity-specific population incidence rates to predict of ovarian cancer. Validation of the model in multi-ethnic populations will require access to additional datasets and will be addressed in future studies. Future studies may also take into account additional risk factors such as family history, which was not available in the UK Biobank. Although rare on a population basis, family history of ovarian cancer confers a substantial increase in risk for individuals (Stratton *et al.*, 1998, Jervis *et al.*, 2014). Finally, we were not able to classify ovarian cancer into subtypes within these data.

Conclusion

A properly calibrated combined clinical and genetic model that is simple and easy to use provides an improvement in risk prediction in the general population of women who would normally be considered to be at average risk. Because general population screening has been unable to show a mortality reduction benefit, the ability to accurately risk-stratify the population is important so that screening options can be targeted to women who need it most. The identification of high-risk women can allow physicians and patients to engage in joint decision-making discussions around the risks and benefits of traditional and novel screening options or risk-reducing salpingo-oophorectomy.

Acknowledgements

We wish to thank Mr Lawrence Whiting for his invaluable expertise in the management of the large data files from the UK Biobank.

This study was fully funded by Genetic Technologies Limited, which had no role in the conceptualisation, design, data analysis, decision to publish or preparation of the article.

Conflicts of interest

G.S.D., N.M.M. and R.A. are employees of Genetic Technologies Limited. E.S. is an employee of Phenogen Sciences Incorporated (a subsidiary of Genetic Technologies Limited). Aspects of this article are covered by Australian Provisional Patent Application No: 2022901988, Methods of assessing risk of developing ovarian cancer.

References

American Cancer Society (2022). Cancer facts & figures 2022. https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2022.html. [Accessed 5 April 2022].

Barnes DR, Rookus MA, McGuffog L, Leslie G, Mooij TM, Dennis J, *et al.*; GEMO Study Collaborators; EMBRACE Collaborators; kConFab Investigators; HEBON Investigators; GENEPSO Investigators; Consortium of Investigators of Modifiers of BRCA and BRCA2 (2020). Polygenic risk scores and breast and epithelial ovarian cancer risks for carriers of BRCA1 and BRCA2 pathogenic variants. *Genet Med* **22**:1653–1666.

Buys SS, Partridge E, Black A, Johnson CC, Lamerato L, Isaacs C, *et al.*; PLCO Project Team (2011). Effect of screening on ovarian cancer mortality: the Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening randomized controlled trial. *JAMA* **305**:2295–2303.

Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, *et al.* (2018) The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**:203–209.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**:7.

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**:520–524.

Dareng EO, Tyrer JP, Barnes DR, Jones MR, Yang X, Aben KKH, *et al.*; GEMO Study Collaborators; GC-HBOC Study Collaborators; EMBRACE Collaborators; OPAL Study Group; AOCS Group; KConFab Investigators; HEBON Investigators; OCAC Consortium; CIMBA Consortium (2022). Polygenic risk modeling for prediction of epithelial ovarian cancer risk. *Eur J Hum Genet* **30**:349–362.

Hanscombe KB, Coleman JRI, Traylor M, Lewis CM (2019). Ukbtools: an R package to manage and query UK Biobank data. *PLoS One* **14**:e0214311.

Hu C, Hart SN, Gnanaolivu R, Huang H, Lee KY, Na J, *et al.* (2021). A population-based study of genes previously implicated in breast cancer. *N Engl J Med* **384**:440–451.

Jervis S, Song H, Lee A, Dicks E, Tyrer J, Harrington P, *et al.* (2014). Ovarian cancer familial relative risks by tumour subtypes and by known ovarian cancer genetic susceptibility variants. *J Med Genet* **51**:108–113.

Koshiyama M, Matsumura N, Konishi I (2017). Subtypes of ovarian cancer and ovarian cancer screening. *Diagnostics (Basel)* **7**:E12.

Kuchenbaecker KB, McGuffog L, Barrowdale D, Lee A, Soucy P, Dennis J *et al.* (2017). Evaluation of polygenic risk scores for breast and ovarian cancer risk prediction in BRCA1 and BRCA2 mutation carriers. *J Natl Cancer Inst* **109**:djw302.

La Vecchia C (2017). Ovarian cancer: epidemiology and risk factors. *Eur J Cancer Prev* **26**:55–62.

Lee A, Yang X, Tyrer J, Gentry-Maharaj A, Ryan A, Mavaddat N *et al.* (2021). Comprehensive epithelial tubo-ovarian cancer risk prediction model incorporating genetic and epidemiological risk factors. *J Med Genet* Published Online First: 29 November 2021. doi: 10.1136/jmedgenet-2021-107904.

Li K, Hüsing A, Fortner RT, Tjønneland 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**:1257–1265.

Liu MC, Oxnard GR, Klein EA, Swanton C, Seiden MV; CCGA Consortium (2020). Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Ann Oncol* **31**:745–759.

McMullen M, Karakasis K, Rottapel R, Oza AM (2021). Advances in ovarian cancer, from biology to treatment. *Nat Cancer* **2**:6–8.

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**:1618–1627.

Menon U, Gentry-Maharaj A, Burnell M, Singh N, Ryan A, Karpinskyj C, *et al.* (2021). Ovarian cancer population screening and mortality after long-term follow-up in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial. *Lancet* **397**:2182–2193.

National Comprehensive Cancer Network (2022). NCCN guidelines. Genetic/familial high-risk assessment: breast, ovarian, and pancreatic. Version 2.2022. https://www.nccn.org/guidelines/guidelines-detail?category=2&id=1503. [Accessed 6 May 2022].

Office for National Statistics (2019). Cancer registration statistics. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/cancerregistrationstatisticscancerregistrationstatisticsengland. [Accessed 20 July 2021].

Office of National Statistics (2016). Mortality statistics. https://www.nomisweb.co.uk/query/construct/summary.asp?mode=construct&version=0&dataset=161. [Accessed 20 July 2021].

Office of National Statistics (2019). Mortality statistics - underlying cause, sex and age. https://www.nomisweb.co.uk/query/construct/summary.asp?mode=construct&version=0&dataset=161. [Accessed 20 July 2021].

Phelan CM, Kuchenbaecker KB, Tyrer JP, Kar SP, Lawrenson K, Winham SJ, *et al*.; AOCS study group; EMBRACE Study; GEMO Study Collaborators; HEBON Study; KConFab Investigators; OPAL study group (2017). Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. *Nat Genet* **49**:680–691.

Purcell S, Chang C (2020). Plink 1.9. https://www.cog-genomics.org/plink/1.9/. [Accessed 30 March 2021].

Rosenthal AN, Fraser LSM, Philpott S, Manchanda R, Burnell M, Badman P, *et al*.; United Kingdom Familial Ovarian Cancer Screening Study collaborators (2017). Evidence of stage shift in women diagnosed with ovarian cancer during phase II of the United Kingdom Familial ovarian cancer screening study. *J Clin Oncol* **35**:1411–1420.

Skates SJ (2012). Ovarian cancer screening: development of the risk of ovarian cancer algorithm (ROCA) and ROCA screening trials. *Int J Gynecol Cancer* **22 Suppl 1**:S24–S26.

Skates SJ, Greene MH, Buys SS, Mai PL, Brown P, Piedmonte M, *et al*. (2017). Early detection of ovarian cancer using the risk of ovarian cancer algorithm with frequent CA125 testing in women at increased familial risk - combined results from two screening trials. *Clin Cancer Res* **23**:3628–3637.

Song H, Ramus SJ, Tyrer J, Bolton KL, Gentry-Maharaj A, Wozniak E, *et al*.; Australian Cancer (Ovarian) Study; Australian Ovarian Cancer Study Group; Ovarian Cancer Association Consortium (2009). A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2. *Nat Genet* **41**:996–1000.

StataCorp (2021). *Stata statistical software: release 17*. StataCorp LLC.

Stratton JF, Pharoah P, Smith SK, Easton D, Ponder BA (1998). A systematic review and meta-analysis of family history and risk of ovarian cancer. *Br J Obstet Gynaecol* **105**:493–499.

Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, *et al*. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* **12**:e1001779.

Yang X, Leslie G, Gentry-Maharaj A, Ryan A, Intermaggio M, Lee A, *et al*. (2018). Evaluation of polygenic risk scores for ovarian cancer risk prediction in a prospective cohort study. *J Med Genet* **55**:546–554.