Risk Prediction of Primary Ovarian Insufficiency in Childhood Cancer Survivors Using Polygenic Risk Scores and Clinical Risk Score

by

Lin Yu

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Epidemiology

School of Public Health

University of Alberta

Abstract

Background

Primary ovarian insufficiency (POI) is one concerning adverse effect of cancer treatments for female childhood cancer survivors. Currently, treatment-related risk factors have been established and used for risk prediction. However, the incremental value of genetic information in the form of a polygenic risk score (PRS) in risk prediction for POI is unknown. We investigated the prediction potential of menopause-related PRSs by integrating them into the existing clinical prediction model.

Methods

A total of 1985 participants in the Childhood Cancer Survivor Study (CCSS) original cohort were used in this study. The published genome-wide association (GWA) studies for age at natural menopause conducted in the general population was used to construct a general population-based PRS (*g*PRS), while top genetic risk associations ($P<10^{-5}$) from a published GWAS of POI among female cancer survivors participating in the St. Jude Lifetime Cohort Study (SJLIFE) was used to evaluate a cancer survivor-based PRS (*c*PRS). The clumping and thresholding (C+T) method was applied to construct additional *c*PRS (named *ct*PRS) for the CCSS samples under more liberal linkage disequilibrium and p-value thresholds. Time-specific logistic regression models were developed for risk prediction. A clinical risk score (CRS), modified from a previous study, was included in the models as an offset term to account for the

clinical risk factors. The added value of the PRSs (*g*PRSs and *c*PRSs) were examined by including the PRS as a main effect in the time-specific logistic regression. The interaction between PRS and ovarian radiation therapy (RT: yes/no) [PRS*RT] and CRS [PRS*CRS] were also evaluated. The area under the ROC curve (AUC), average precision (AP), scaled Brier Score (sBrS), Spiegelhalter-z statistic, and the calibration curve were computed using a 5-fold cross-validation framework to assess the model performance. Finally, these metrics were compared with those of the baseline clinical prediction model – CRS model.

Results

Sixteen PRSs were constructed in total, including three *g*PRSs (PRS48, PRS69, and PRS262) computed from two GWA studies for age at natural menopause in the general population, thirteen *c*PRSs (PRS6, and additional twelve *ct*PRSs developed from C+T method) constructed from a childhood cancer survivor-based GWA study for POI. The AUC, AP, and sBrS value of the baseline -- CRS model were 0.797 (95% CI: 0.778, 0.816), 0.539 (95% CI: 0.502, 0.574), and 0.236 (95% CI: 0.203, 0.267), respectively. The main effect models with any PRSs performed similarly: the AUC values were between 0.775 to 0.780, the AP values ranged from 0.530 to 0.532, indicating adequate performance but no improvement in the discrimination compared to the baseline model; the Spiegelhalter-z statistics decreased from 11.427 to a range between 0.099 and 0.154, implying improved calibration of PRS main effect models; the overall performance, i.e., sBrS, slightly improved compared to the CRS model, ranging from 0.236 to 0.27. The AUC, AP and sBrS estimates of the PRS*CRS models were similar to those of the CRS model, ranged from 0.792 to 0. 799, 0.528 to 0.537, and 0.227 to 0.257, respectively. The Spiegelhalter-z

statistics, ranging from 3.774 to 6.686, decreased compared to that of the CRS model. The PRS*RT models performed similarly to the PRS main effect models: the AUCs were in a range of 0.775 -0.780; the AP ranged from 0.528 to 0.537; the Spiegelhalter-z statistics ranged from 1.401 to 1.485. The calibration curve showed that the CRS model underestimated the risk of POI. And the PRS main effect and the PRS*RT models performed similarly: the models overestimated the risk for participants in high-risk groups (actual risk >0.5) and underestimated the risk for low-risk groups (actual risk <0.5). The PRS*CRS models calibrated well initially while overestimating the risk with the observed risk increase. None of the PRSs in the main effect models was significant, while both PRS*CRS and PRS*RT interactions were statistically significant in the interaction models.

Conclusions

Incorporating the genetic information in the predictive model did not improve the discrimination but sometimes improved the model calibration. In summary, the *g*PRS main effect model and *ct*PRS*RT interaction models had similar performance and were the best models among all: the overall performance improved compared to the CRS model, where the improvement came from better model calibration. The generalizability of the models should be assessed in external validation using external data in the future.

Preface

This thesis is an original work by Lin Yu. The research projects, of which this thesis is apart, received research ethics approval from the University of Alberta Research Ethics Board, project names “Epidemiological analysis of clinical and genetics data”, Pro00085976 approved on October 15, 2018 , and “Risk prediction for ovarian failure in childhood cancer survivors” Pro00067066, approved on August 19, 2016.

Acknowledgments

I would like to express my gratitude and appreciation to everyone who has supported me through the process of my thesis research. In particular, I would like to give a huge thank you to Dr. Yan Yuan, my supervisor and mentor who has always been there to guide me throughout my entire graduate journey. I would like to thank her for her continuous support and assistance in getting me through the difficult times. I would not have made it this far without her!

I would like to thank my co-supervisor, Dr. Cindy Im, for providing me with valuable insight and advice on my thesis research directions and feedback on my thesis writing. I would like to thank all the members in Dr. Yuan's research group for helping me. I want to acknowledge and thank Zhe Lu, in particular, for his advice and for taking the time to help me!

I would like to thank the Canadian Centre for Applied Research in Cancer Control (ARCC) for supporting me through the ARCC Studentship, and offering me the opportunity to share my research at the ARCC conference. I would like to acknowledge and thank Dr. Russell J. Brooke for providing the necessary GWAS data of the SJLIFE cohort for my research. I would also like to thank the Canadian Institutes for Health Research for supporting this project.

I would like to acknowledge and thank the Childhood Cancer Survivor Study for providing the data for this research. The study participants and their families for participating in the CCSS

study and completing the surveys. I would like to thank the researchers who made the GWAS data publicly available.

I would like to thank the School of Public Health at the University of Alberta for providing me with this opportunity to learn and grow. I would like to thank the professors and staff who have assisted me in addressing my questions throughout my graduate degree.

Finally, I would like to thank my parents and family for supporting me. I would like to express my gratitude for their support of my decision to study abroad. I would like to thank my friends in Edmonton for always treating me like a family member and keeping me company and my friends in China for constantly encouraging me.

Table of Contents

List of Figures

List of Tables

List of Appendices

List of Appendix Figures

List of Appendix Tables

List of Abbreviations

AP	Averaged precision
AUC	Area under the ROC curve
CCSS	Childhood Cancer Survivor Study
***c*PRS**	Cancer survivor-based PRS
CRS	Clinical risk score
***ct*PRS**	Clumping and thresholding-based PRS using survivor GWAS and CCSS
FPR	*False-positive rate*
***g*PRS**	General population-based PRS
GWAS	Genome-wide association study
IPCW	Inverse-probability-of-censoring weighting
LD	linkage disequilibrium
NOBOX	Newborn ovary homeobox gene
POI	Primary ovarian insufficiency
PR	Precision-recall
PRS	Polygenic risk score
ROC	The receiver operating characteristic curve
RT	Radiotherapy
sBrS	scaled Brier Score
SJLIFE	St. Jude Lifetime Cohort Study
SNP	Single nucleotide polymorphism
TPR	True-positive rate

1 Introduction

Childhood cancer survivors are a rapidly growing group in developed countries due to the advancement in cancer treatments(1). In the late 1980s, 71% of children diagnosed with cancer will survive at least five years after their initial cancer diagnosis. With the improvement in cancer treatment, the five-year survival rate is over 80% in North America(2). This is a significant improvement, and consequently, the size of the childhood cancer survivor population has grown dramatically, to approximately 300,000 individuals in Canada and 483,000 survivors in the US(2,3). However, the remarkable increases in survival have been accompanied with adverse effects later in life known as late effects due to the toxicity of cancer treatments(4). Approximately two-thirds of childhood cancer survivors experience late effects, which may include cardiopulmonary, endocrine, renal or hepatic dysfunction, reproductive difficulties, neurocognitive impairment, psychosocial difficulties and the development of subsequent cancers(5).

One of the late effects for female cancer survivors is primary ovarian insufficiency (POI), characterized by permanent natural cessation of menstruation before 40 years old(6). POI can occur early, during, or immediately following the completion of cancer treatment(7) or, more commonly, in the years that follow the completion of cancer treatments prior to age 40(8). In the general population, the prevalence of POI is approximately 1%(9), whereas a study reported that 10.9% (100 out of 921) of childhood cancer survivors developed POI(8). Additionally, another

study reported a 13-fold (95% CI = 3.26 to 53.51; P<.001) increased risk in developing POI in childhood cancer survivors compared to their non-survivor siblings(10).

Risk prediction models for POI have been proposed to identify individuals at high risk of developing POI. Treatment-related risk factors, such as radiation therapy and chemotherapy, have been studied and incorporated into a predictive model for POI in female survivors of childhood cancer(11,12). In recent decades, GWA studies have identified genetic variants associated with menopause-related phenotypes(13,14). However, little is known regarding the potential of menopause-related polygenic risk score (PRS) to identify POI at different risk levels, and the effect of PRS-treatment interactions on POI. Therefore, this study aims to develop prediction models using genetic information from GWA studies in combination with clinical risk factors, and investigate potential PRS-treatment interactions.

The purpose of the predictive models is to identify individuals who are at high risk of POI. For individuals at high risk of developing POI, patients can be counseled before or shortly after cancer treatment regarding the risk and the need for fertility preservation such as oocyte and ovarian tissue cryopreservation(15). If the risk of developing POI is low, clinicians can provide counseling to patients and their families to avoid suffering and cost to undergo interventions for fertility preservation. By providing quantitative evidence for potential POI risk among cancer survivors, the ultimate goal is to improve the quality of life among female childhood cancer survivors.

This thesis is structured as follows. The remainder of Chapter 1 reviews the literature on POI, previous related risk modeling work, and genetic susceptibility for POI. Chapter 1 also describes the statistical methods used in model development and evaluation, followed by an overview of the data for model development. Chapter 2 presents the construction of PRS using genetic data. Chapter 3 highlights the development and assessment of risk prediction models incorporating PRS. Finally, Chapter 4 summarizes the findings, discusses study limitations, and provides recommendations for future research.

1.1 Literature Review

The Prevalence of Primary Ovarian Insufficiency

It is estimated that approximately 90% of women experience menopause between the age of 45 and 55 years(16), with the median age at natural menopause occurring between 50 and 52 years of age(17). *Primary ovarian insufficiency* (POI) occurs if a woman experiences menopause naturally before age 40, and the prevalence of POI is about 1% in the general population(18). However, among childhood cancer survivors, the continually growing five-year survival rate (exceeds 80% already) adds to an increasing prevalence of POI as a result of cancer therapies; currently, about 10% of cancer survivors experience POI(19).

Impact of Primary Ovarian Insufficiency

Individuals with POI are more likely to develop chronic diseases. *Shah et al.* have shown that patients have an increased probability of developing cardiovascular disease in the post-menopause years(20). Lower estrogen levels following menopause also increase the risk of developing hypertension and ventricular remodeling(21). At the same time, menopause and chronic diseases also place a mental strain on both patients and their families. It has been shown that women with ovarian dysfunction were more likely to develop anxiety and depression (22–24).

A primary concern of POI is infertility, as fertility is irreversible after POI onset. Some fertility preservation options, such as oocyte and ovarian tissue cryopreservation, are available to preserve reproductive function(9,25,26). However, a study has suggested that cancer patients feel challenged to make decisions about fertility preservation(27). One reason is that these decisions are time-pressured. Many participants reported that they did not have enough time for decision-making before cancer treatment(28). Additionally, uncertainty

makes it challenging to make decisions(29). For example, women have to trade-off the risk of developing POI after cancer treatments with no guarantee of favorable outcomes from the fertility preservation procedure. Some decision aid methods have been proposed(30), which mainly focused on improving patients' knowledge about fertility preservation.

Clinical Risk Factors and Clinical Risk Prediction Model

Extensive studies have been undertaken to identify treatment-related risk factors associated with compromised reproductive function following cancer treatment(31). Radiation exposure to the ovary, abdominal or pelvic can induce genomic damage in oocytes and the surrounding granulosa cells, leading to either a decreased or exhausted ovarian follicle pool depending on the extent of the damage(32). Also, chemotherapy agents, especially alkylating agents (such as busulfan, cyclophosphamide, lomustine, procarbazine, etc.) can prevent cell division and growth by interacting with DNA and reduce the number of follicles for maturation and reproduction, increasing the risk for ovarian dysfunction(33).

Well-established clinical risk factors and cancer treatments mentioned before have been evaluated as risk factors to develop risk prediction models for compromised reproductive function in female cancer survivors. For example, *Clark et al.* has used clinical predictors such as cumulative alkylating drug dose(11), radiation exposure to the ovary, abdomen and pelvis, age at cancer diagnosis, and hematopoietic stem-cell transplant receipt to build models to predict an individual's risk for developing menopause within five years following a cancer diagnosis(11). *Lu* has investigated the association between clinical risk factors and their potential interactions on POI and incorporated them into a predictive model for developing POI risk at different ages in female cancer survivors(12).

Genetic Risk for Primary Ovarian Insufficiency

Aside from clinical risk factors, genetic studies have shown that POI is a complex, heterogeneous disorder that is influenced by multiple genetic components(34). Genetic studies have shown that the genetic variations on the X chromosome contribute mostly to the etiology of POI. Meanwhile, increasing attention has focused on autosome single-gene variations known to regulate follicle development and maturation. For example, a 2009 study discovered that the Newborn ovary homeobox gene (NOBOX) plays a critical role in early folliculogenesis(35). Deficiency in NOBOX disrupts early folliculogenesis and oocyte-specific gene expression, accelerating post-natal oocyte loss and abolishing the transition from primordial to growing follicles. Moreover, the mutations in follicle-stimulating hormone receptor genes, such as FSHR and LHCGR, were found to affect the development and maturation of follicles and oocytes(36).

Many techniques have been applied to discover the genetic variants associated with different phenotypes. A popular technique, *Genome-wide Association Study* (GWAS), has been widely conducted to study the genetics of phenotypes in recent decades(13), which aims to find common variants associated with phenotypes(37). The GWAS technique consists of screening and comparing genetic variants in patients with the disease of interest and healthy controls. In GWAS, the association between a genetic variant to the phenotype is observed when there is a statistically significant difference in the allelic frequency of the genetic variant between diseased patients and healthy controls after controlling for multiple testing.

Potential of Polygenic Risk Score

Although GWA studies have led to the identification of many variants associated with POI, the effect size of each genetic variant identified from GWAS is typically small and accounts for only a small fraction of heritability, meaning that single variants have limited predictive power(38). Thus, I evaluated the genetic risk

for POI in the form of a PRS, or a score that combines the estimated effects of many disease-/phenotype-associated genetic variants reported in published GWAS data. PRS has been proposed as a genetic risk prediction tool for a wide range of diseases. According to the PGS catalog, an open database of published polygenic risk scores, 829 PRSs were built for 214 traits as of July 2021(39). A clinically useful PRS would allow clinicians to identify individuals at elevated risk of disease, thus informing disease screening(40), therapeutic interventions(41), and life planning(42) to prevent or delay disease onset.

1.2 Methodology Review

1.2.1 Polygenic Risk Score

A polygenic risk score (PRS, also termed a *polygenic score* or *genetic risk score*) estimates an individual's genetic liability to a trait or disease(38). An individual's PRS is a sum of genotypes at selected genetic loci, weighted by corresponding genotype effect size estimates derived from GWAS summary statistics. The *base data* (discovery GWA studies), consisting of summary statistics of the association between genetic variants and phenotypes, is required to construct a PRS. Then the PRS can be calculated for the individuals in the *target data*, which usually includes new samples independent of the base samples with genotype and phenotype data. Finally, the potential of the PRS in risk prediction can be examined using the target data. The target data was introduced in the following section (Section 1.3).

The general steps for computing a PRS include: *a)* selecting base data; *b)* quality control of the base and target data; *c)* and finally, the construction of PRS. The workflow for constructing a PRS is presented in **Figure 1-1**. The rest of this section included more details.

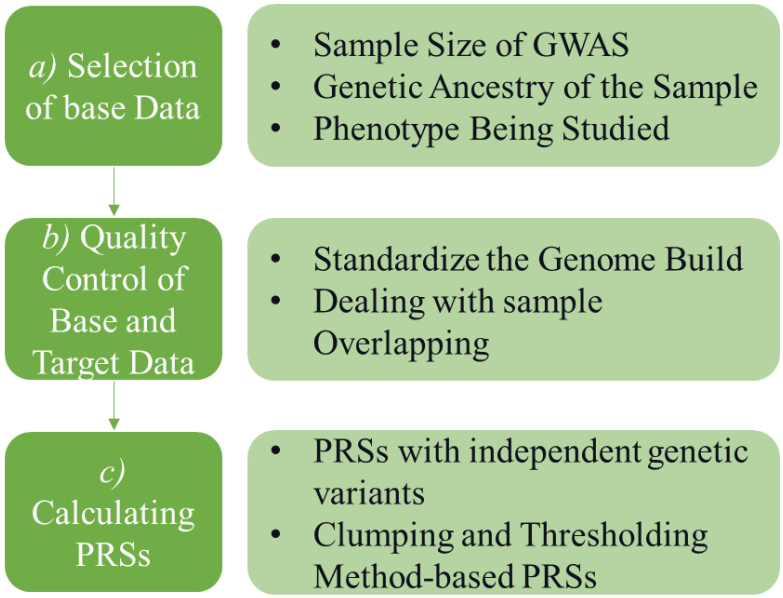


Figure 1-1 The workflow for PRS construction

a) *Selection of Base Data (GWA studies)*

When it comes to the selection of GWA studies, I considered several factors as listed below.

Sample Size of GWAS: Given that the PRS is built based on the summary statistics from GWA studies, the GWA studies should have a reasonable sample size so that it is powerful enough to detect the association between genetic variants and phenotype and ultimately increase the predictive ability of the PRS(38,43). GWA studies with a relatively larger sample size are recommended.

Genetic Ancestry of the Sample: Allele frequencies and correlations between genetic variants (i.e., linkage disequilibrium patterns) vary among different ancestry groups. The estimated effect size of a given genetic variant might be different from population to population. For this reason, the ancestry of the base and target samples should be the same to ensure an accurate estimation.

Phenotype Being Studied: The phenotype studied in the base data should be the same as the phenotype of interest in the PRS construction. However, some phenotypes might share similar genetic architecture(44). GWA

studies for other traits that are relevant to the phenotype of interest could be a substitute when GWA studies for phenotypes of interest are not available

b) Quality Control of Base and Target Data

Following the base data selection is the quality control, mainly including standardizing the genome build and removing overlapping samples between the base and target data.

Standardize the Genome Build: Currently, there are several different genome builds(45). If the positions of the genetic variants in the base and target data differ by genome build, genomic positions between the base and target data should be standardized(46). Online tools such as Ensembl(47) or LiftOver (48) can be used to standardize the genome build across base and target data.

Sample Overlapping: Study samples that overlap between the base and target data must be removed as sample overlapping can result in inflation of the association between the PRS and the phenotype tested in the target data(49).

c) Calculating Polygenic Risk Score

The next step is to calculate the PRS for individuals in the target data. An individual's PRS is usually a weighted sum of the number of risk alleles and thus can be given by:

$$PRS = \sum_i \beta_i x_i \qquad \text{Equation 1-1}$$

Where i is an indicator for the ith genetic variant, β_i is the log odds ratio or the coefficient of linear regression in the base data, and x_i is the number of risk alleles at the ith genetic variant. (note: risk allele refers to the allele of a genetic variant that associated with the disease risk)

I considered the following two strategies for the PRS construction:

1) *PRS with independent genetic variants:* A common way to construct a PRS is to utilize the independent genetic variants that are genome-wide significant (P-value $<5x10^{-8}$). The advantage of this method is that the target data is not necessary.

2) *Clumping and Thresholding Method:* The clumping and thresholding (C+T) method(50–52) is another popular way to derive PRS. Specifically, the clumping(C) step forms a clump around the most significant genetic variant (index genetic variant). This clump included the nearby variants within some genetic distance (for example, within 250 kilobases) of the index genetic variant. Then the algorithm computes the pairwise correlation or linkage disequilibrium (or LD R^2) between the index genetic variant and the nearby variants using the reference samples' genotyping data. All nearby variants that are correlated with this index variant beyond a pre-specified LD R^2 threshold (e.g., 0.1, 0.4, etc.) are removed. The clumping step goes on with the next most significant genetic variant (new index variant) that has not been removed yet. In summary, clumping iteratively circles through genetic variants so that only variants with LD R^2 less than a pre-specified value are kept. The thresholding (T) step removes the remaining genetic variants with GWAS association P-value under a certain threshold (e.g., P $<5x10^{-5}$). The C+T method provides more flexibility in comparison to the abovementioned strategy 1). Researchers can specify different LD R^2 and P-value thresholds. Target data is used to decide which combination of LD R^2 and P-value threshold allows the PRS to perform best in the risk prediction.

1.2.2 Risk Prediction

Time-specific Logistic Regression

Logistic regression is one of the most commonly used approaches for modeling the relationship between variables and a binary outcome. The general form of logistic regression can be expressed as:

$$ln(p/(1-p)) = \beta 0 + \beta 1\ X1 + \beta 2\ X2 \ldots + \beta kXk \qquad \text{Equation 1-2}$$

where p denotes the probability of the outcome of interest occurring, $p/(1-p)$ represents the odds of the outcome, and ln denotes the natural logarithm. $X1$, $X2$, ... Xk are the predictors (could be potential interactions) and $\beta 1$, $\beta 2$, …, βk are their coefficients, and $\beta 0$ is the intercept. Logistic regression models the log odds of a binary outcome using a linear combination of predictors. Several assumptions are required to use logistic regression, including a linear relationship between the continuous covariates and the log odds of the outcome, no multicollinearity among the covariates, and independence among individuals.

Logistic regression explores the risk factors for the outcome and determines the relative importance of risk factors with respect to the outcome. The odds ratio, given by exp(β), is used to measure the association. For a continuous variable, the odds ratio can be interpreted as holding all other covariates constant, the odds of the outcome of interest occurring will change by exp(β) for one unit increase in the covariate value. For a categorical variable, the odds ratio can be interpreted as: the odds of the outcome of interest occurring in a specific category is exp(β) times the odds of that in the reference category.

Logistic regression is very popular for its interpretability. However, logistic regression cannot handle *time-to-event* data, as logistic regression does not take the time of the event's occurrence into consideration in the modeling. In this study, POI is determined by two elements: the ovarian status and the age at menopause (i.e.,

time-to-event data). Thus, censoring is a concern when modeling POI risk. For example, a survivor's POI status was censored if she was 27 at her last follow-up and had a normal menstrual function. She was still at risk of POI. However, the actual outcome is not observable due to censoring.

The inverse-probability-of-censoring weighting (IPCW) method was employed in previous studies to account for the censored observations(53). Briefly, the IPCW weights were obtained by modeling the censoring process using the same set of covariates, such as age at diagnosis and radiation dosage to the ovary, for modeling the POI status(12). Censored individuals will thus contribute to the risk model through the IPCW weights. Individuals with known POI status will be given weights in the estimation of the logistic regression model. Therefore, we call this model "time-specific logistic regression".

The previous risk prediction study used the multiple imputation technique to deal with the missing data(54). In the multiple imputation procedure, each missing value will obtain multiple imputed values given the observed data. As illustrated in the hypothetical example in **Figure 1-2**, patients 2 and 3 have one missing value in variable 1 (Var 1) and variable 2 (Var 2), respectively. The multiple imputation technique assigned five values for the missing values. The Multiple imputation has been shown to yield unbiased parameter estimates when the assumption of missing at random is satisfied. It also reduces information waste, which is common in complete case analysis. The previous risk prediction study has imputed each missing value with five values(12). The imputed values were used in this study. Therefore, the target data consists of five copies of the original CCSS data.

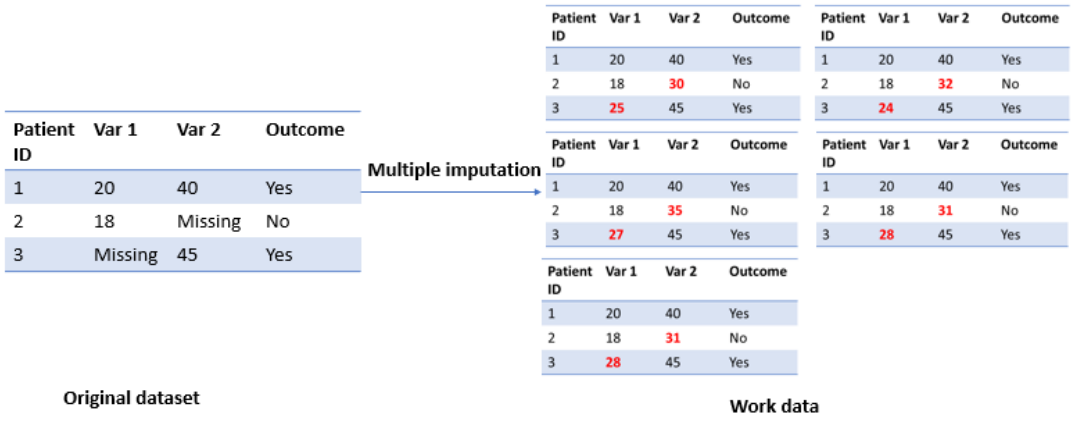


Figure 1-2 A hypothetical example of multiple imputation

Apart from censoring and missing, the competing risk event of surgical premature menopause (SPM, had bilateral oophorectomy) needs to be considered. Female childhood cancer survivors who had bilateral oophorectomy before the age of 40 would not develop menopause naturally (i.e., are no longer at risk of natural menopause). The competing risk is considered in the IPCW method, where there is an indicator variable for the event (menopause, menstruation, or surgical premature menopause)

1.2.3 Evaluation of Model Performance

Five-fold Cross-Validation

The internal validation was performed under a five-fold cross-validation framework. **Figure 1-3** shows the process of cross-validation. In step 1, the eligible observations in the dataset were randomly divided into five roughly equal-sized parts. In step 2, One part of the data (noted as D1) was left out as the validation set. The remaining four parts were combined as the training set (D2-D5, blue part). The risk prediction model was developed using the training set and applied to obtain the predicted risk for patients in the validation set (D1). This process was repeated as shown in step 3; thus, each observation in the work data gets a predicted risk. As we have five copies of work data (multiple imputation was applied to deal with missing data), the procedure in

Figure 1-3 was repeated over the five datasets. Therefore, each observation will have five predicted risks. Finally, the five predicted risks were averaged to represent an individual's risk, and the averaged risk was compared to the observed POI status to assess the model performance mentioned below.

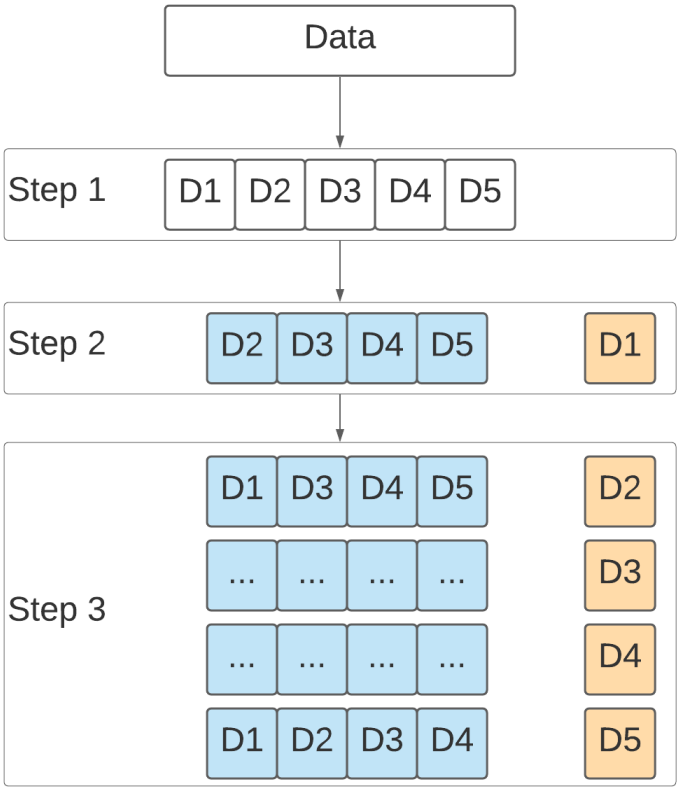


Figure 1-3 The workflow for the cross-validation framework

We need to evaluate the quality of the predictive model after model development. For the prediction of a binary outcome, we typically examine three different perspectives:

1. the closeness of the predictions to the observed outcome (overall performance),
2. the discrimination power of the prediction model, and
3. the calibration of the prediction model.

The following performance measures are used and will be calculated under a five-fold cross-validation framework.

The Overall Performance: Scaled Brier Score

The overall performance is quantified by the distance between the predicted outcome (for a binary outcome, it refers to the predicted probability) and the observed outcome. The squared difference, defined as the Brier score, is usually calculated to quantify the distance(55). The formula of Brier score is calculated as the mean squared error of the prediction, written as:

$$BS = \frac{1}{N}\sum_{i=1}^{N}(p_i - y_i)^2 \qquad \text{Equation 1-3}$$

Where p_i is the probability of the outcome of the *i*th individual, *N* is the number of individuals being predicted, y_i is the actual observed outcome of the *i*th individual. Brier score corresponds to the goodness of fit of the model, and it takes values in the interval [0,1]. A smaller distance between the predicted and observed outcome indicates a better prediction.

However, the range of Brier scores changes by incidence rate. For instance, with incidence = 0.5, the Brier score will range from 0 for a perfect prediction to a maximum of 0.25 for a non-informative model where the overall proportion of the event of interest is usually given as the predicted probability of the event occurring for each individual in the dataset. While with incidence decreased from 0.5 to 0.1, the maximum Brier score will decrease to 0.09. Thus, the interpretation and comparison of multiple Brier scores is challenging when Brier scores are calculated from different incidence rates.

One alternative way to evaluate the overall performance is to use *scaled Brier Score (sBrS)*.(56). It is calculated as:

$$sBrS = 1 - \frac{BS}{BS_{ref}} = (\frac{BS_{ref} - BS}{BS_{ref}}) \quad \text{Equation 1-4}$$

where BS_{ref} is the Brier score of the baseline prediction model(s) which need to be improved. The baseline prediction models could be any pre-existing models. More commonly, a null model is employed as a baseline model, in which each individual in the data set is given a constant predicted probability, which is the overall proportion of the outcome of interest in the study sample. The Brier score for a null model is calculated as:

$$BS_{ref} = \frac{1}{N}\sum_{i=1}^{N}(\bar{y} - p_i)^2 \quad \text{Equation 1-5}$$

Where $\bar{y}$ is the event rate of the outcome of interest of the study samples, and p_i is the probability of the outcome of the *i* th individual, and *N* is the number of individuals in the work data.

As we can see from equation 1-4, $sBrS$ is very similar to the coefficient of determination (R^2) in the linear regression(57). $sBrS$ measures the fractional amount of improvement in the Brier score of a model compared to the baseline model and is more interpretable than the Brier score. A $sBrS$ of zero indicates that the prediction is only as good as the baseline model, and a $sBrS$ of one suggests that the prediction is perfect. A $sBrS$ less than zero indicates that the prediction is even worse than that of the null model(58).

The Discrimination

Discrimination assesses a predictive model's ability to discriminate those who have the outcome from those who do not, i.e., the model's ability to accurately classify the outcome as event or non-event. The area under the receiver operating characteristic (ROC) curve is the most commonly used metric to evaluate the discrimination of prediction algorithms. The ROC curve plots *true-positive rate* (TPR, or *sensitivity*) versus *false-positive rate* (FPR, or 1-*specificity*) of the predictive models over all possible cutoffs for the probability of an outcome, where true positive rate and false positive rate were calculated as:

The area under the ROC curve (AUC) summarizes the model performance over all possible cutoffs. AUC represents the probability that a randomly chosen observation with a positive outcome (case) has a higher predicted risk score than a randomly chosen observation with a negative outcome (noncase). According to the definition of AUC, it can be calculated by randomly selecting a case and a non-case as a pair, then calculating the proportion of pairs where the case has a higher predicted risk score than that of the non-case over all possible pairs.

AUC ranges from 0.5 to 1.0. An AUC of 0.5 (A ROC curve with the line at 45 degrees) represents the true positive rate equals the false positive rate, indicating that the predictive model is unable to discriminate between positive and negative outcomes. An AUC value greater than 0.5 implies that the prediction model has some ability to distinguish between positive and negative outcomes. The higher the AUC, the better the discrimination performance of the predictive algorithm. A perfect predictive model has an AUC of 1.0.

The Averaged Precision (AP)

The precision-recall (PR) curve is developed to evaluate the prediction model. The PR curve plots the positive predicted value (also termed as precision) against the sensitivity (also termed as recall). The area under the PR curve, also known as *averaged precision*, is a measurement summarizing the precision over all possible thresholds(59). the AP is shown more appropriate for the outcome of interest of lower prevalence in the target population than AUC, independent of the event rate. The AP has a value between the event rate of the outcome of interest and 1.0, with AP = event rate representing a noninformative model, and a bigger AP indicating better predictive power of the prediction model.

The Calibration

Though the discrimination (usually quantified by AUC) is popularly used, a high AUC value does not indicate accurate risk prediction as AUC is a ranking metric(60). For example, for an algorithm with adequate discrimination, the risk estimates for all samples may be systematically overly high, regardless of whether they developed the outcome or not. Therefore, the agreement between the observed outcomes and predicted values needs to be assessed to evaluate the reliability of the predictions. The agreement, often called calibration, is another key property of predictive models though it is far less reported(61). Some guidelines for prediction modeling studies, such as the TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) guideline recommends reporting on the calibration performance(62). The calibration is more clinically useful than discrimination when the algorithm is used for informing decision-making. A poorly calibrated risk estimate may lead to a false decision-making. Take the POI risk prediction model as an example, it is unacceptable to under- or over-estimate the risk of developing POI. For example, if a patient's POI risk is underestimated, she may opt out of the fertility preservation treatment, thus denying her the opportunity to have a biological child. For a patient with an over-estimated POI risk, the fertility preservation would expose her to unnecessary risk as these procedures are usually invasive, and lead to possible harmful complications.

Therefore, the under- and over-estimate may lead to under- and over-treatment. Reporting the calibration performance could possibly prevent incorrect and potentially harmful clinical decisions. A well-calibrated model is that for individuals with a predicted risk of p%, the observed frequency of events among those individuals should be approximately p%.(58). For example, if the mean predicted probability for a group of individuals is 20%, the model is well-calibrated if the observed probability is close to 20%.

Spiegelhalter-z Statistic

The Brier score is an overall measure of performance and can be decomposed to calibration and discrimination. Spiegelhalter-z statistic is used to measure the calibration aspect of the Brier score(63). By expanding the square in equation 1-5, the Brier score can be decomposed into:

$$BS = \frac{1}{N}\sum_{i=1}^{N}(y_i - p_i)^2$$
$$= \frac{1}{N}\sum_{i=1}^{N}(y_i - p_i)(1 - 2p_i) + \frac{1}{N}\sum_{i=1}^{N} p_i(1 - p_i) \qquad \text{Equation 1-6}$$

The first term measures the lack of calibration, and the second term measures the lack of discrimination of the predictions. The Spiegelhalter-z statistic is defined as:

$$z = \frac{\sum_{i=1}^{n}(y_i - p_i)(1 - 2p_i)}{\sum_{i=1}^{n}\sqrt{(1 - 2p_i)^2 p_i(1 - p_i)}} \qquad \text{Equation 1-7}$$

It asymptotically follows a standard normal distribution. The Spiegelhalter-z statistic has a value of zero under the null hypothesis of perfectly calibrated, statistically significant scores. (i.e., $\boldsymbol{z} < -1.96$ or $\boldsymbol{z} > 1.96$) generally indicate poor calibration.

Calibration Plot

The plot helps visualize the calibration. The calibration plot usually has the prediction on the x-axis and the outcome on the y-axis. The perfect predictions should be at the 45-degree line and help for orientation. For binary outcomes, we can first group individuals by similar predicted probabilities, e.g., group by quantiles. Then we compare the mean observed proportions with the event of each subgroup to its mean predicted probabilities. Moreover, smoothing techniques such as the loess algorithm can be applied to estimate the relationship between the observed proportions of the event and the predicted probabilities.

1.3 The CCSS Original Cohort

The Childhood Cancer Survivor Study (CCSS), a multi-institutional retrospective cohort study, was established to prospectively assess late mortality, subsequent neoplasms, adverse cardiac and pulmonary outcomes, fertility outcomes, and health-related behaviors of cancer survivors(64,65). The CCSS original cohort was first established in 1994 and recruited survivors in North America that met the following eligibility criteria: (1) diagnosed between January 1st 1970 and December 31st 1986; (2) age less than 21 at diagnosis; (3) lived for at least five years after the date of a cancer diagnosis.

1.3.1 Baseline Demographic Data

A baseline questionnaire was completed by 14,054 eligible subjects in CCSS original cohort. The questionnaire consisted of 289 questions regarding demographics, medical care practices, prescription medications, medical conditions, and so on. Five follow-up surveys were released to obtain updated information regarding health conditions and monitor potential adverse effects of cancer treatment. Demographic information was requested from CCSS. Of particular interest to this study are age, sex, race, cancer type, age at cancer diagnosis(66).

1.3.2 Treatment Data

All participants who completed the baseline questionnaire were asked to sign a consent form to allow access to all medical records since their cancer diagnosis. Then a detailed summary of cancer treatment information, including chemotherapy, radiation therapy, and surgery information for each cohort member, was extracted from the medical record. Data was collected regarding the specific chemotherapy agents and their respective doses and radiation dose and location. A list of treatment variables is provided in Appendix A.

1.3.3 Genotyping Data

The buccal cell genomic DNA information of eligible participants who had completed the baseline questionnaire was also collected during May 1999 and June 2006(67). A specimen collection kit was mailed to participants, which included a cover letter outlining the study, a consent form, an instruction sheet, a 45 mL bottle of mouthwash, a specimen collection container, return mail labels, and postage. The participants were asked to return the sample to the Molecular Genetic Laboratory in Cincinnati, Ohio. A total of 5739 participants have genotype data available, among which 2958 are female participants(68). The genotyping data was requested from CCSS, and standard quality control (QC) of CCSS original cohort genotype data was performed(69).

Definition of POI Outcomes

The outcome of interest in this study is POI, defined as either: (1) experiencing menopause naturally before the age of 40 years, or (2) never experiencing menarche by age 18(6). Two variables are needed to determine the outcomes: Ovarian status and the age at the event onset. The above-established definition and patients' self-reported menstrual history information in the baseline and follow-up 1, 4, and 5 questionnaires help classify

patients as POI, surgical premature menopause (SPM, had bilateral oophorectomy), or normal. Ambiguous cases whose ovarian status could not be determined were further manually reviewed by endocrinologists, based on the patients' responses for menstrual history questions; age at event onset is derived from the CCSS surveys. The age at last menstrual period or surgical time informs age at menopause or surgical menopause, respectively. For patients with incomplete age information, age at event onset was imputed in consultation with endocrinologists.

CCSS Study Sample Eligibility

Data of the female cancer survivors in the CCSS original cohort study was obtained in order to develop risk predictive models for POI in female childhood cancer survivors. The inclusion and exclusion criteria were applied according to previous studies(12). The detailed exclusion criteria are included in Appendix B. Results from explanatory data analysis are provided in Appendix C.

1.4 References

1. Lund L, Schmiegelow K, Rechnitzer C, Johansen C. A Systematic Review of Studies on Psychosocial Late Effects of Childhood Cancer: Structures of Society and Methodological Pitfalls May Challenge the Conclusions. Pediatric blood & cancer. 2011 Apr 1;56:532–43.

2. Canada PHA of. Cancer in Children in Canada (0-14 years) [Internet]. 2012 [cited 2021 Jun 30]. Available from: https://www.canada.ca/en/public-health/services/chronic-diseases/cancer/cancer-children-canada-0-14-years.html

3. Stats - kids cancer care [Internet]. [cited 2021 Aug 6]. Available from: https://www.kidscancercare.ab.ca/childhood-cancer/stats

4. Marwick C. Childhood cancer survivors experience long term side effects. BMJ. 2003 Sep 6;327(7414):522.

5. Nelson LM. Primary Ovarian Insufficiency. N Engl J Med. 2009 Feb 5;360(6):606–14.

6. Gelson E, Prakash A, Macdougall J, Williams D. Reproductive health in female survivors of childhood cancer. The Obstetrician & Gynaecologist. 2016;18(4):315–22.

7. Levine JM, Whitton JA, Ginsberg JP, Green DM, Leisenring WM, Stovall M, et al. Nonsurgical premature menopause and reproductive implications in survivors of childhood cancer: A report from the Childhood Cancer Survivor Study: NSPM and Reproductive Implications. Cancer. 2018 Mar 1;124(5):1044–52.

8. Chemaitilly W, Li Z, Krasin MJ, Brooke RJ, Wilson CL, Green DM, et al. Premature ovarian insufficiency in childhood cancer survivors: a report from the St. Jude Lifetime Cohort. The Journal of Clinical Endocrinology & Metabolism. 2017;102(7):2242–50.

9. Rudnicka E, Kruszewska J, Klicka K, Kowalczyk J, Grymowicz M, Skórska J, et al. Premature ovarian insufficiency – aetiopathology, epidemiology, and diagnostic evaluation. Prz Menopauzalny. 2018 Sep;17(3):105–8.

10. Sklar CA, Mertens AC, Mitby P, Whitton J, Stovall M, Kasper C, et al. Premature Menopause in Survivors of Childhood Cancer: A Report From the Childhood Cancer Survivor Study. J Natl Cancer Inst. 2006 Jul 5;98(13):890–6.

11. Clark RA, Mostoufi-Moab S, Yasui Y, Vu NK, Sklar CA, Motan T, et al. Predicting acute ovarian failure in female survivors of childhood cancer: a cohort study in the Childhood Cancer Survivor Study (CCSS) and the St Jude Lifetime Cohort (SJLIFE). The Lancet Oncology. 2020 Mar 1;21(3):436–45.

12. Lu Z. Risk Prediction for Premature Ovarian Insufficiency in Childhood Cancer Survivors. University of Alberta; 2020.

13. Perry JRB, Corre T, Esko T, Chasman DI, Fischer K, Franceschini N, et al. A genome-wide association study of early menopause and the combined impact of identified variants. HumMolGenet. 2013;22(7):1465–72.

14. He C, Kraft P, Chen C, Buring JE, Paré G, Hankinson SE, et al. Genome-wide association studies identify loci associated with age at menarche and age at natural menopause. Nat Genet. 2009 Jun;41(6):724–8.

15. Baker VL. Primary ovarian insufficiency in the adolescent. Current Opinion in Obstetrics and Gynecology. 2013 Oct;25(5):375–81.

16. Miro F, Parker SW, Aspinall LJ, Coley J, Perry PW, Ellis JE. Sequential classification of endocrine stages during reproductive aging in women: the FREEDOM study. Menopause. 2005 Jun;12(3):281–90.

17. Gold EB, Bromberger J, Crawford S, Samuels S, Greendale GA, Harlow SD, et al. Factors associated with age at natural menopause in a multiethnic sample of midlife women. American Journal of Epidemiology. 2001 May 1;153(9):865–74.

18. Haller-Kikkatalo K, Uibo R, Kurg A, Salumets A. The prevalence and phenotypic characteristics of spontaneous premature ovarian failure: a general population registry-based study. Human reproduction. 2015;30(5):1229–38.

19. Younis JS. Ovarian aging and implications for fertility female health. Minerva Endocrinol. 2012;37(1):41–57.

20. Shah D, Nagarajan N. Premature menopause – Meeting the needs. Post Reprod Health. 2014 Jun 1;20(2):62–8.

21. Zhao Z, Wang H, Jessup JA, Lindsey SH, Chappell MC, Groban L. Role of estrogen in diastolic dysfunction. American Journal of Physiology-Heart and Circulatory Physiology. 2014;306(5):H628–40.

22. Faubion SS, Kuhle CL, Shuster LT, Rocca WA. Long-term health consequences of premature or early menopause and considerations for management. Climacteric. 2015 Jul 4;18(4):483–91.

23. Choi IY, Choi YE, Nam HR, Lee JW, Park EC, Jang SI. Relationship between Early Menopause and Mental Health Problems. Korean Journal of Family Practice. 2018 Feb 20;8(1):87–92.

24. Lawson AK, Klock SC, Pavone ME, Hirshfeld-Cytron J, Smith KN, Kazer RR. Prospective study of depression and anxiety in female fertility preservation and infertility patients. Fertil Steril. 2014 Nov;102(5):1377–84.

25. Cabry R, Merviel P, Hazout A, Belloc S, Dalleac A, Copin H, et al. Management of infertility in women over 40. Maturitas. 2014 May 1;78(1):17–21.

26. Chian R-C, Quinn P. Fertility Cryopreservation. Cambridge University Press; 2010. 287 p.

27. Jones G, Hughes J, Mahmoodi N, Smith E, Skull J, Ledger W. What factors hinder the decision-making process for women with cancer and contemplating fertility preservation treatment? Hum Reprod Update. 2017 Jul 1;23(4):433–57.

28. Kirkman M, Winship I, Stern C, Neil S, Mann GB, Fisher JRW. Women's reflections on fertility and motherhood after breast cancer and its treatment. Eur J Cancer Care (Engl). 2014 Jul;23(4):502–13.

29. Gonçalves V, Sehovic I, Quinn G. Childbearing attitudes and decisions of young breast cancer survivors: a systematic review. Hum Reprod Update. 2014 Apr;20(2):279–92.

30. Mahmoodi N, Bekker H, King N, Hughes J, Jones G. Decision Aids' Efficacy to Support Women's Fertility Preservation Choices Before Cancer Treatment: An Environmental Scan. In: ISDM 2017 Abstract Book: Oral communications. Leeds; 2017.

31. Green DM, Sklar CA, Boice Jr JD, Mulvihill JJ, Whitton JA, Stovall M, et al. Ovarian failure and reproductive outcomes after childhood cancer treatment: results from the Childhood Cancer Survivor Study. Journal of Clinical Oncology. 2009;27(14):2374.

32. Adriaens I, Smitz J, Jacquet P. The current knowledge on radiosensitivity of ovarian follicle development stages. Human Reproduction Update. 2009 May 1;15(3):359–77.

33. Laven JSE. Genetics of Early and Normal Menopause. SeminReprodMed. 2015;33(6):377–83.

34. Pu D, Xing Y, Gao Y, Gu L, Wu J. Gene variation and premature ovarian failure: a meta-analysis. Eur J Obstet Gynecol Reprod Biol. 2014 Nov;182:226–37.

35. van Dooren MF, Bertoli-Avellab AM, Oldenburg RA. Premature ovarian failure and gene polymorphisms. CurrOpinObstetGynecol. 2009;21(4):313–7.

36. Bosch E, Alviggi C, Lispi M, Conforti A, Hanyaloglu AC, Chuderland D, et al. Reduced FSH and LH action: implications for medically assisted reproduction. Human Reproduction. 2021 Jun 1;36(6):1469–80.

37. Christin-Maitre S, Tachdjian G. Genome-wide association study and premature ovarian failure. Annales d Endocrinologie. 2010;71(3):218–21.

38. Choi SW, Mak TS-H, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. Nat Protoc. 2020 Sep;15(9):2759–72.

39. Lambert SA, Gil L, Jupp S, Ritchie SC, Xu Y, Buniello A, et al. The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. Nat Genet. 2021 Apr;53(4):420–5.

40. Seibert T, Fan C, Wang Y, Zuber V, Karunamuni R, Parsons J, et al. Polygenic hazard score to guide screening for aggressive prostate cancer: Development and validation in large scale cohorts. BMJ. 2018 Jan 10;360:j5757.

41. Collins FS, Varmus H. A new initiative on precision medicine. New England journal of medicine. 2015;372(9):793–5.

42. Khera AV, Emdin CA, Drake I, Natarajan P, Bick AG, Cook NR, et al. Genetic risk, adherence to a healthy lifestyle, and coronary disease. New England Journal of Medicine. 2016;375(24):2349–58.

43. Dudbridge F. Power and predictive accuracy of polygenic risk scores. PLoS Genet. 2013 Mar;9(3):e1003348.

44. Louwers YV, Visser JA. Shared Genetics Between Age at Menopause, Early Menopause, POI and Other Traits. Frontiers in Genetics. 2021;12:1889.

45. Genome browser FAQ [Internet]. [cited 2021 Aug 13]. Available from: https://genome.ucsc.edu/FAQ/FAQreleases.html#release1

46. Data changes that occur between builds [Internet]. National Center for Biotechnology Information (US); 2005 [cited 2021 Aug 13]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK44467/

47. Ensembl genome browser 104 [Internet]. [cited 2021 Aug 13]. Available from: https://uswest.ensembl.org/index.html

48. Lift genome annotations [Internet]. [cited 2021 Aug 13]. Available from: http://genome.ucsc.edu/cgi-bin/hgLiftOver

49. Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, Visscher PM. Pitfalls of predicting complex traits from SNPs. Nat Rev Genet. 2013 Jul;14(7):507–15.

50. Privé F, Vilhjálmsson BJ, Aschard H, Blum MGB. Making the Most of Clumping and Thresholding for Polygenic Scores. The American Journal of Human Genetics. 2019 Dec 5;105(6):1213–21.

51. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. Genome Medicine. 2020 May 18;12(1):44.

52. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations | Nature Genetics [Internet]. [cited 2021 Aug 10]. Available from: https://www.nature.com/articles/s41588-018-0183-z

53. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests. Biometrics. 2000 Sep;56(3):779–88.

54. Rubin DB. Multiple imputation for nonresponse in surveys. Vol. 81. John Wiley & Sons; 2004.

55. Monthly Weather Review. War Department, Office of the Chief Signal Officer; 1950. 418 p.

56. Glahn HR, Jorgensen DL. CLIMATOLOGICAL ASPECTS OF THE BRIER P-SCORE. Monthly Weather Review. 1970 Feb 1;98(2):136–41.

57. Hu B, Palta M, Shao J. Properties of R2 statistics for logistic regression. Statistics in Medicine. 2006;25(8):1383–95.

58. Stevens RJ, Poppe KK. Validation of clinical prediction models: what does the “calibration slope” really measure? Journal of Clinical Epidemiology. 2020 Feb 1;118:93–9.

59. Yuan Y, Su W, Zhu M. Threshold-Free Measures for Assessing the Performance of Medical Screening Tests. Frontiers in Public Health. 2015;3:57.

60. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, Bossuyt P, et al. Calibration: the Achilles heel of predictive analytics. BMC Medicine. 2019 Dec 16;17(1):230.

61. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. Journal of clinical epidemiology. 2016;74:167–76.

62. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. 2015 Jan 6;162(1):W1-73.

63. Spiegelhalter DJ. Probabilistic prediction in patient management and clinical trials. Statistics in Medicine. 1986;5(5):421–33.

64. Robison LL, Armstrong GT, Boice JD, Chow EJ, Davies SM, Donaldson SS, et al. The Childhood Cancer Survivor Study: A National Cancer Institute–Supported Resource for Outcome and Intervention Research. J Clin Oncol. 2009 May 10;27(14):2308–18.

65. Robison LL. The Childhood Cancer Survivor Study: a resource for research of long-term outcomes among adult survivors of childhood cancer. Minn Med. 2005 Apr;88(4):45–9.

66. Childhood Cancer Survivor Study - 2013 Progress Report [Internet]. Childhood Cancer Survivor Study; 2013 [cited 2020 Sep 20]. Available from: https://ccss.stjude.org/content/dam/en_US/shared/ccss/documents/progress-report-2013.pdf

67. Buccal Cell Collection [Internet]. St. Jude CCSS. [cited 2021 Oct 31]. Available from: https://ccss.stjude.org/biospecimens/buccal-cell-collection.html

68. Public access GWAS data tables [Internet]. [cited 2021 Aug 7]. Available from: https://ccss.stjude.org/develop-a-study/gwas-data-resource/public-access-gwas-data-tables.html

69. Biospecimens | St. Jude CCSS [Internet]. [cited 2021 Oct 16]. Available from: https://ccss.stjude.org/biospecimens.html

2 Polygenic Risk Score Construction

2.1 Introduction

GWA studies have been used to identify genetic variants associated with diseases. They have revealed that most conditions have a polygenic pattern where many genetic variants – individually having only a small effect on the disease – contribute to the development of disease(1). A PRS quantifies the cumulative effect of many genetic variants for a particular disease(2). The most typical method for calculating a PRS is to add up the number of risk alleles (0, 1, or 2) that an individual carries for each genetic locus included in the PRS, weighted by the effect sizes of the risk alleles reported by a GWAS for the disease of interest. The weights are usually given as beta coefficients (effect sizes) from linear regression or log odds ratios for binary outcomes. The significance (P-value) of the genetic variants in the GWA studies is often used to determine whether a genetic variant should be included in the PRS (3).

GWA studies have also been done for menopause-related phenotypes. For POI, there have been *five* GWA studies so far. Each of the five studies had a small sample size, with the largest including roughly 1300 people. Four of them were performed among people of European or East Asian ancestry in the general population. One study was conducted among childhood cancer survivors (4), where 799 female cancer survivors in the St. Jude Lifetime Cohort Study (SJLIFE) were used to identify genetic variants associated with clinically diagnosed POI. In the analysis of female survivors, models were adjusted for alkylating agents (cyclophosphamide equivalent dose) and ovarian radiotherapy (RT) doses. The results showed that 20 out of 830,884 genotyped variants had a P value less than 10^{-5}. Among them, 13 genetic variants (upstream of the Neuropeptide Receptor 2 gene) were associated with POI prevalence. A haplotype formed by 4 of the 13 variants showed association

with POI for patients exposed to ovarian RT, indicating an interaction between genetics and radiation treatment. Replication was performed in 1624 survivors from the Childhood Cancer Survivor Study (CCSS), and the association was also observed between the haplotype and POI for patients exposed to ovarian radiotherapy (OR= 3.97, 95%CI= [1.67 to 9.41], P= .002).

Aside from GWA studies for POI, thirteen GWA studies conducted in the general population with larger study sample sizes are available for age at natural menopause, the biggest of which is a meta-analysis of GWA studies conducted in the general population with up to 69,360 women of European ancestry(5). The study showed that 1,208 out of a total of ~2.6 million genetic variants reached the genome-wide significance threshold ($P < 5 \times 10^{-8}$) for association with age at natural menopause. Among those significant genetic variants, 54 independent genetic variants at 44 genetic loci were identified, explaining 6% of the variance in age at natural menopause. The variance explained increased to 21% for the top 29,958 independent variants with an association P value less than 0.05. A more recent study included 119,160 samples (European ancestry) from the UK Biobank and examined about 9.5 million genetic variants. This larger study identified 74 independent genetic variants that were associated with POI(6). A summary of these studies is provided in Appendix E.

Ideally, a GWAS for POI would be the best data source for studying POI polygenic risk prediction. However, the general population's existing GWA studies for POI are limited by their small sample sizes. Therefore, I focused on the GWA studies for a highly related complex trait, age at natural menopause, to formulate POI-related polygenic risk scores. It has been hypothesized that age at natural menopause and POI are possible manifestations of the same underlying genetic susceptibility owing to the inheritance patterns observed(7). Studies have also shown that POI and age at natural menopause share common genetic factors involved in DNA repair and maintenance(8).

Although the success of GWAS has successfully identified genetic variants associated with human diseases, the small effect size of a single variant on disease has fundamentally limited the ability to use a single genetic variant in disease risk prediction and disease diagnosis(9,10). In 2009, a study used multiple genetic variants to evaluate the risk of individuals developing schizophrenia, demonstrating the ability of common variants in disease risk assessment(11). Since then, a number of studies have used multiple genetic variants to assess disease risk: the most notable is a 2018 study demonstrating the use of multiple-gene predictors in the form of PRS to stratify individuals' risk for five common diseases(12). As of July 2021, over 800 PRSs have been published for more than 200 different traits(13).

To our knowledge, PRSs have not been developed and used in risk prediction for POI in childhood cancer survivors or the general population. Therefore, we propose to evaluate genetic risk using PRSs based on the summary statistics from GWASs conducted in both the general population and childhood cancer survivors. The evaluation of the general population PRS for POI in survivors may not only have clinical utility for the survivor population but can also provide further insights into the relative contribution of general population PRS to POI in childhood cancer survivors. Furthermore, PRSs can be used to investigate gene-by-environment interactions and risk stratification(14).

2.2 Methods

Base Data

As discussed in Section 1.2.1-a), base data (i.e., GWA studies summary statistics) is needed to compute a PRS. GWA studies for age at natural menopause conducted in the general population were pulled from two public databases: GWAS catalog(15) and GWASatlas(6). The summary statistics of a GWA study conducted in the

female childhood cancer survivors using the SJLIFE cohort were obtained by corresponding with the study's original authors (4). Summary statistics, including genetic variant identifiers, genome build, effect allele, reference allele (non-effect allele), effect allele frequency, regression coefficient (effect size), standard error (of the regression coefficient), sample size, and P-value, were extracted.

Target Data

The study samples in the Childhood Cancer Survivor Study (CCSS) original cohort (introduced in Section 1.3) were used as the target data. The summary statistics in the base data were used to calculate the PRS for all individuals in the target data. For the C+T method, aside from base data, target data is also used to compute the LD R^2 required in the clumping step.

Inclusion and exclusion criteria

The inclusion and exclusion criteria for base data (GWA studies) were discussed carefully in Section 1.2.1-a). and can be summarized as follows:

- GWA studies conducted in general population or cancer survivor sample(s) of predominantly European ancestry;
- GWA studies with relatively larger sample size;
- Phenotype definition in GWA studies/meta-analyses is POI or is relevant for the study of menopause-related phenotypes (for example, age at natural menopause);
- GWA studies conducted with appropriate standard sample/variant quality control procedures
- GWA studies with summary statistics data, including:

- **Genetic variant identifiers**: the coordinate of the genetic variant including chromosome and base pair position, human genome assembly/build
- **Effect allele**: the allele that was coded for association testing and can either increase or decrease risk.
- **Reference allele**: the non-effect allele
- **Effect allele frequency**: the frequency of the effect allele of a genetic variant in the population
- **Regression coefficient/effect size**: the change in the trait value or disease risk with each additional copy of the effect allele (i.e., the coefficients of regression)
- **Standard error**: standard error of the regression coefficient
- **Sample size**: the sample size of the base data
- **P-value**: the significance of the genetic variant

The inclusion and exclusion criteria for the target data (CCSS original cohort) have been established in a previous study (Appendix B). Furthermore, Participants without genotype data and who were not of European genetic ancestry were excluded. Individuals who overlapped with the SJLIFE were also excluded from the study as target data should be independent of the base data to avoid potential inflation. (discussed in Section 1.2.1-b)).

Quality Control for Base and Target Data

Quality control for the base and data was conducted before computing PRS: 1) the Ensembl online coordinate converting tool(16) was used to standardize the genetic variant coordinates to the hg19 genome build. GWA studies with missing reference allele information were imputed. Given the provided dbSNP identifiers, the missing reference allele information was imputed from the cited dbSNP reference (dbSNP build 129), which

was obtained from the UCSC Genome Browser(17); 2) the genetic variants that were matched in both the base and target data were kept for analysis (detailed matching algorithm is given in Appendix D).

Constructing the PRS

Two strategies were applied for the PRS construction: 1) only genetic variants that achieved genome-wide significance in their respective GWA studies were included; or 2) the C+T method was applied to the cancer survivor-based GWA study to build several *c*PRSs (named *ct*PRSs) (discussed in Section 1.2.1-c)). For the C+T method, the LD R^2 between genetic variants was computed using the CCSS samples' genotyping data. Different LD R^2 and P-value thresholds were considered to compute several candidate PRSs: LD R^2 = {0.1,0.4,0.8}; P-value = {$5x10^{-2}$, $5x10^{-3}$, $5x10^{-4}$, $5x10^{-5}$}. For LD calculations, genomic windows of 250 kilobases were used. The PRSs were calculated using PLINK (version 1.9)(18) and R (version 4.0.3) (https://cran.r-project.org/bin/windows/base/)(19).

2.3 Results

Selection of Base Data

In total, eighteen GWA studies were examined. A summary of the eighteen GWA studies is available in Appendix E. After applying the inclusion and exclusion criteria, the two most recent GWA studies conducted in the general population for age at natural menopause were selected for computing general population-based PRSs (*g*PRSs). The two GWA studies identified 74 and 54 independent genetic variants associated with age at natural menopause among 9.5 and 2.6 million genetic variants, respectively. The only GWAS for POI performed among childhood cancer survivors in the SJLIFE cohort was used to construct a cancer survivor population-based PRS (*c*PRS) despite the limited sample size.

Table 2-1 Summary of the relevant information of the three selected GWA studies.

Phenotype:	**Age at natural menopause**		**POI**
Study	Watanabe *et al.*, 2019	Day *et al*., 2015	Brooke *et al*., 2019
Meta-analysis component	yes	yes	No
Nhit*	74	54	None
Discovery cohort	119160	38,968 European women	799 female cancer survivors with 85.7% of European ancestry and 14.3% of African ancestry
Replication cohort	NA	14,435 European women	1624 female cancer survivors with 98.3% of European ancestry and 1.7 of African ancestry
Number of quality-controlled genetic variants	~ 9.5 million	~ 2.6 million	~ 830 thousand
Phenotype measurement	Self-report	Self-report	Self-report with clinical assessment
Phenotype definition	Age at last menstrual period	Age at last menstrual period	Primary ovarian insufficiency (Yes/No)

Note: Nhit is the number of independent genetic variants associated with the phenotype using $P<5x10^{-8}$ as the critical value.

Target data

Of the 4541 female childhood cancer survivors, 4432 female survivors remained in the study sample after applying the exclusion criteria established by previous studies. Additionally, 1977 individuals were excluded for missing genotype data, 250 individuals not of European descent were excluded, and 220 individuals involved in the SJLIFE POI GWAS were excluded. The total study sample consisted of 1985 individuals.

The Polygenic Risk Score

As discussed in the methods, two strategies were applied for the PRS construction. In total, sixteen PRSs, including three *g*PRSs and thirteen *c*PRSs. The results are summarized in Figure 2-1.

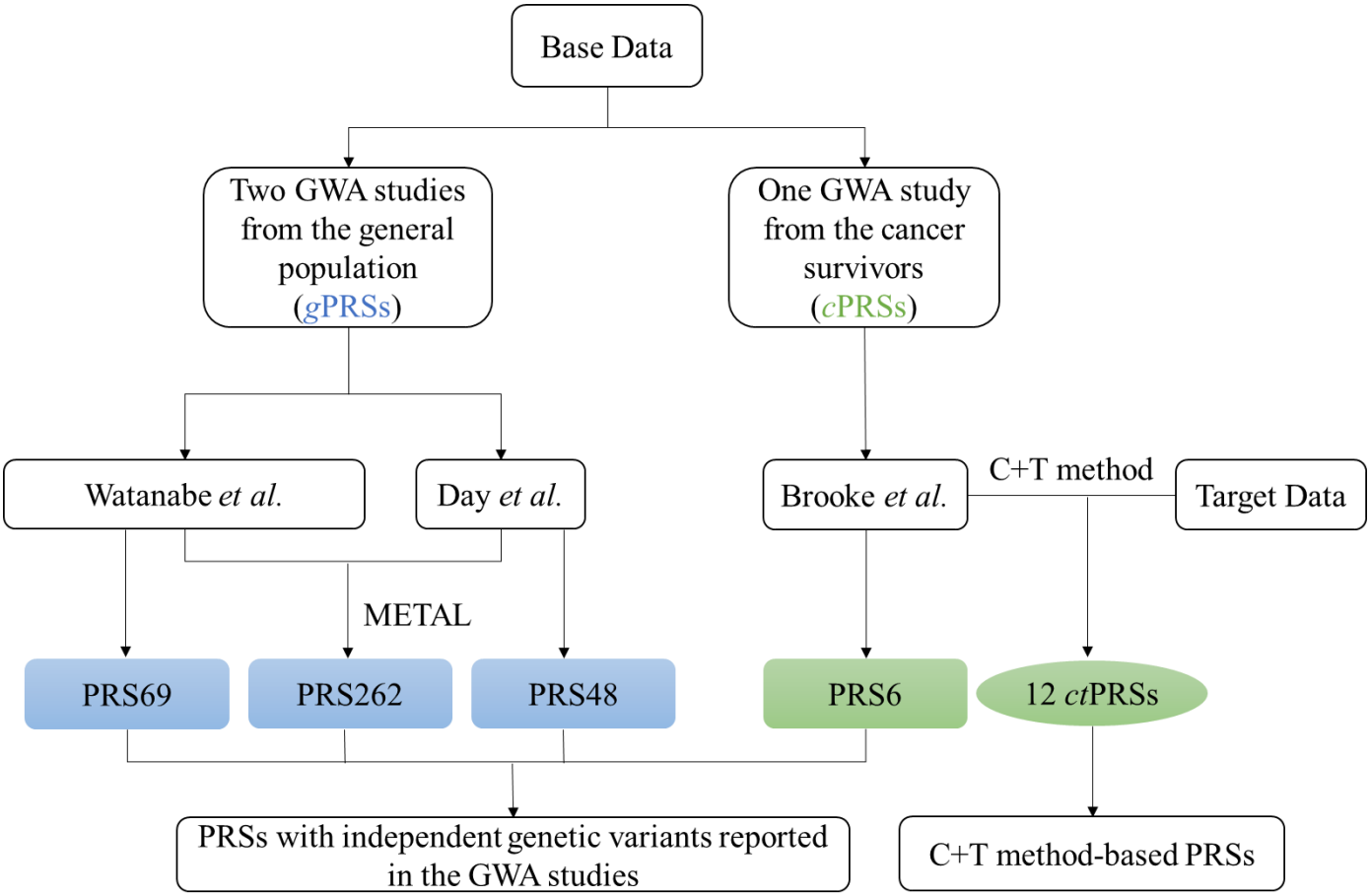
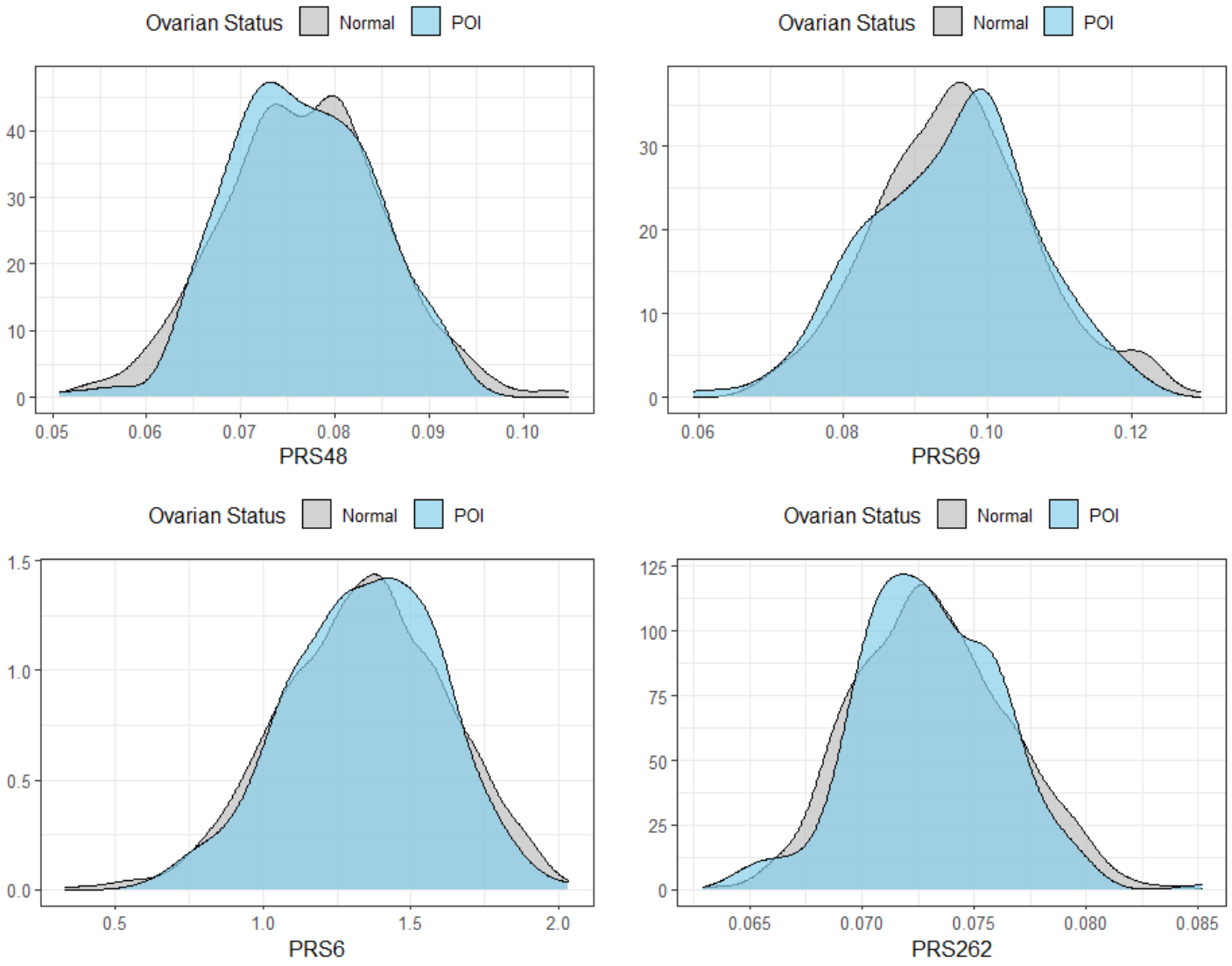


Figure 2-1 An overview of the PRS construction

1) PRSs with independent genetic variants

The independent significant genetic variants reported in the three selected GWA studies were used directly to build the PRS, represented by PRS48, PRS69, and PRS6, where the numbers in the names refer to the number of genetic variants that were included in that PRS. METAL was applied to meta-analyze results from the two GWASs conducted in the general population(20) (the METAL methodology is described in Appendix G). The

Genome-wide significant ($P<5x10^{-8}$) genetic variants with LD R^2 less than 0.1 were kept, resulting in 262 independent genetic variants, which were further used to build a new PRS, named PRS262. The density plots of the four PRSs are given in Figure 2-2.



Note: the x-axis is the PRS, and the y-axis is the density

Figure 2-2 Density plots of gPRSs/*c*PRS by ovarian status

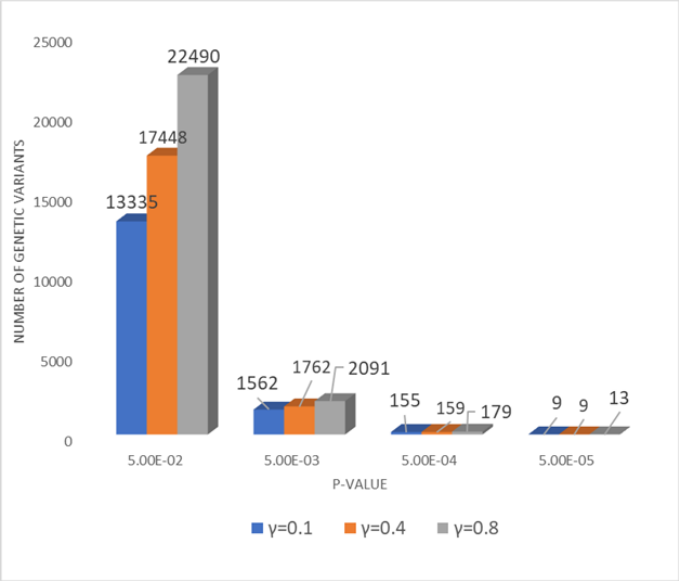
Overall speaking, the distributions of PRS among POI cases overlapped with that among normal individuals. It was seen that for PRS48 and PRS6, the frequency of more extreme PRS is larger in the normal group; however,

for PRS69 and PRS262, the frequency of PRS with extremely low value is relatively higher in the POI cases than that in the normal group.

2) C+T method-based PRSs

The C+T method-based PRSs (*ct*PRSs) were computed using the summary statistics from the cancer survivor-based GWA study and CCSS samples. Note: *the x-axis represents the P-value threshold; the y-axis represents the number of genetic variants selected under each setting, and γ represents the pairwise correlation threshold (LD R2) between genetic variants.*

Figure 2-3 shows the number of genetic variants that were selected by the C+T method with different parameter settings. As expected, a larger correlation coefficient and P-value allow more genetic variants to be chosen. Appendix H shows the distribution of *ct*PRSs in different parameter settings.



Note: the x-axis represents the P-value threshold; the y-axis represents the number of genetic variants selected under each setting, and γ represents the pairwise correlation threshold (LD R^2) between genetic variants.

Figure 2-3 The number of genetic variants selected under 12 different hyperparameter settings

2.4 Discussion

Combining common genetic variants into a polygenic risk score has been shown to identify individuals at a high level of disease risk successfully. Such stratification could inform disease screening, therapeutic interventions, and life planning to prevent or delay disease onset.

Three general population-based PRSs (PRS48, PRS69, PRS262) were computed from two GWA studies conducted in the general population for age at natural menopause. One female cancer survivor-based PRS (PRS6) was constructed given the GWAS summary statistics obtained from the female childhood cancer survivors for POI, plus twelve candidate *ct*PRSs were computed using the cancer survivor GWAS and CCSS samples. The density plots showed that the PRS distributions between the POI cases and controls were similar.

One challenge in building *c*PRSs is the limited availability and sample sizes of GWA studies among childhood cancer survivors. In my research, the *c*PRSs (PRS6 and 12 *ct*PRSs) were constructed using the only existing POI GWAS in the childhood cancer survivors. Same as the *g*PRSs, the development of the *c*PRSd can be improved, and validation can be done when more extensive GWA studies in childhood cancer survivors become available.

The C+T method is easy to apply and interpret compared to alternative methods. However, there are two concerns when applying the C+T method. First, in this study, the target data – CCSS original cohort was used as the reference panel to calculate the LD R^2. The reference panel should be selected carefully. The LD R^2 estimates could be biased if the study samples of the reference panel and the samples in the initial GWAS study came from different populations. Second, the C+T method used the original effect estimates from the GWA

studies in the PRS construction. However, better PRSs may be built using more advanced techniques which can re-estimate the effect sizes of genetic variants.

2.5 References

1. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. Genome Medicine. 2020 May 18;12(1):44.

2. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. Genome research. 2007;17(10):1520–8.

3. Konuma T, Okada Y. Statistical genetics and polygenic risk score for precision medicine. Inflammation and Regeneration. 2021 Jun 17;41(1):18.

4. Brooke RJ, Chemaitilly W, Wilson CL, Krasin MJ, Li Z, Im C, et al. A high-risk genetic profile for premature menopause (PM) in childhood cancer survivors (CCS) exposed to gonadotoxic therapy: A report from the St. Jude Lifetime Cohort (SJLIFE) and Childhood Cancer Survivor Study (CCSS). JCO. 2017 May 20;35(15):10502–10502.

5. Day FR, Ruth KS, Thompson DJ, Lunetta KL, Pervjakova N, Chasman DI, et al. Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. Nat Genet. 2015 Nov;47(11):1294–303.

6. Genome wide association study ATLAS [Internet]. [cited 2021 Oct 28]. Available from: https://atlas.ctglab.nl/traitDB/3366

7. Rossetti R, Ferrari I, Bonomi M, Persani L. Genetics of primary ovarian insufficiency. Clin Genet. 2017 Feb;91(2):183–98.

8. Laven JSE. Genetics of Early and Normal Menopause. SeminReprodMed. 2015;33(6):377–83.

9. Dudbridge F. Power and predictive accuracy of polygenic risk scores. PLoS Genet. 2013 Mar;9(3):e1003348.

10. Dudbridge F. Polygenic Epidemiology. Genet Epidemiol. 2016 May;40(4):268–72.

11. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009 Aug;460(7256):748–52.

12. Khera A, Chaffin M, Aragam K, Haas M, Roselli C, Choi S, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nature Genetics. 2018 Sep 1;50.

13. Lambert SA, Gil L, Jupp S, Ritchie SC, Xu Y, Buniello A, et al. The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. Nat Genet. 2021 Apr;53(4):420–5.

14. Mullins N, Power RA, Fisher HL, Hanscombe KB, Euesden J, Iniesta R, et al. Polygenic interactions with environmental adversity in the aetiology of major depressive disorder. Psychological medicine. 2016;46(4):759–70.

15. GWAS catalog [Internet]. [cited 2021 Aug 11]. Available from: https://www.ebi.ac.uk/gwas/

16. Ensembl genome browser 104 [Internet]. [cited 2021 Aug 13]. Available from: https://uswest.ensembl.org/index.html

17. Index of /goldenPath/hg18/database [Internet]. [cited 2021 Aug 11]. Available from: http://hgdownload.soe.ucsc.edu/goldenPath/hg18/database/

18. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. The American journal of human genetics. 2007;81(3):559–75.

19. Ihaka R, Gentleman R. R: a language for data analysis and graphics. Journal of computational and graphical statistics. 1996;5(3):299–314.

20. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26(17):2190–1.

21. Lu Z. Risk Prediction for Premature Ovarian Insufficiency in Childhood Cancer Survivors. University of Alberta; 2020.

22. Robison LL, Armstrong GT, Boice JD, Chow EJ, Davies SM, Donaldson SS, et al. The Childhood Cancer Survivor Study: A National Cancer Institute–Supported Resource for Outcome and Intervention Research. J Clin Oncol. 2009 May 10;27(14):2308–18.

23. Steyerberg EW. Clinical prediction models. Springer; 2019.

24. Boqué R, Rius FX, Massart DL. Straight line calibration: something more than slopes, intercepts, and correlation coefficients. Journal of chemical education. 1994;71(3):230.

25. Zhu X, Tang H, Risch N. Admixture mapping and the role of population structure for localizing disease genes. Adv Genet. 2008;60:547–69.

26. Jin J, Cerise JE, Kang SJ, Yoon EJ, Yoon S, Mendell NR, et al. Principal components ancestry adjustment for Genetic Analysis Workshop 17 data. BMC Proc. 2011 Nov 29;5(Suppl 9):S66.

27. Menozzi P, Piazza A, Cavalli-Sforza L. Synthetic Maps of Human Gene Frequencies in Europeans. Science. 1978;201(4358):786–92.

28. Ma J, Amos C. Theoretical Formulation of Principal Components Analysis to Detect and Correct for Population Stratification. PloS one. 2010 Sep 17;5.

29. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006 Aug;38(8):904–9.

30. Refaeilzadeh P, Tang L, Liu H. Cross-validation. Encyclopedia of database systems. 2009;5:532–8.

31. Using AUC and accuracy in evaluating learning algorithms | IEEE Journals & Magazine | IEEE Xplore [Internet]. [cited 2021 Nov 2]. Available from: https://ieeexplore.ieee.org/abstract/document/1388242?casa_token=QICMJ3G6dSIAAAAA:R0E2GX677ELL6O2Vc-Y5zJCCJKe2QL3pjwbB3X5swqAAXKnnqfIIKaTwKvF55-vRZpOVSxiCXQ

32. Yuan Y, Su W, Zhu M. Threshold-Free Measures for Assessing the Performance of Medical Screening Tests. Frontiers in Public Health. 2015;3:57.

33. Spiegelhalter DJ. Probabilistic prediction in patient management and clinical trials. Statistics in Medicine. 1986;5(5):421–33.

34. Huang J, Ling CX. Using AUC and accuracy in evaluating learning algorithms. IEEE Transactions on knowledge and Data Engineering. 2005;17(3):299–310.

35. Green DM, Sklar CA, Boice Jr JD, Mulvihill JJ, Whitton JA, Stovall M, et al. Ovarian failure and reproductive outcomes after childhood cancer treatment: results from the Childhood Cancer Survivor Study. Journal of Clinical Oncology. 2009;27(14):2374.

36. Mahmoodi N, Bekker H, King N, Hughes J, Jones G. Decision Aids' Efficacy to Support Women's Fertility Preservation Choices Before Cancer Treatment: An Environmental Scan. In: ISDM 2017 Abstract Book: Oral communications. Leeds; 2017.

37. Bzdok D, Engemann D, Thirion B. Inference and Prediction Diverge in Biomedicine. Patterns (N Y). 2020 Oct 8;1(8):100119.

38. Choi SW, Mak TS-H, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. Nat Protoc. 2020 Sep;15(9):2759–72.

39. Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M. Genome-wide association studies in diverse populations. Nature Reviews Genetics. 2010;11(5):356–66.

40. Hickey PF, Bahlo M. X chromosome association testing in genome wide association studies. Genetic epidemiology. 2011;35(7):664–70.

41. Coughlin SS. Recall bias in epidemiologic studies. Journal of clinical epidemiology. 1990;43(1):87–91.

42. Leung K-M, Elashoff RM, Afifi AA. Censoring issues in survival analysis. Annual review of public health. 1997;18(1):83–104.

43. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests. Biometrics. 2000 Sep;56(3):779–88.

44. Wallace WHB, Smith AG, Kelsey TW, Edgar AE, Anderson RA. Fertility preservation for girls and young women with cancer: population-based validation of criteria for ovarian tissue cryopreservation. Lancet Oncol. 2014 Sep;15(10):1129–36.

3 Risk Prediction for Primary Ovarian Insufficiency in Female Childhood Cancer Survivors

3.1 Introduction

The risk prediction model for POI in female childhood cancer survivors has been built mainly using clinical information obtained from clinical records, including demographic information, age at diagnosis, race, bone marrow transplant (Yes/No), total body irritation dose, abdomen radiation dose, minimal ovary radiation dose, and exposures to any of 21 chemotherapy agents (listed in Appendix A) (1) Aside from clinical risks, the utilization of genetic information in the form of PRS in risk stratification for several diseases has been demonstrated in studies in recent years. Using PRS, researchers were able to identify 8.0, 6.1, 3.5, 3.2, and 1.5 percent of the population as having a threefold elevated risk of coronary artery disease, atrial fibrillation, type 2 diabetes, inflammatory bowel disease, and breast cancer, respectively(2).

While the prediction models for POI using treatment exposures as predictors have been established(1), little is known regarding the added value of PRSs and PRS-treatment interactions on top of pre-existing clinical prediction models. Therefore, I evaluated the potential incremental value of the PRSs constructed from the GWAS summary statistics from the general population and childhood cancer survivor studies in improving risk prediction for POI, after accounting for clinical risks.

3.2 Methods

Study Population

The study samples included 1985 female childhood cancer survivors in the CCSS original cohort study (see Sections 1.3)(3).

Statistical Analysis

Time-specific logistic regression at age 40 was used to estimate the POI risk in this study. As mentioned in Section 1.1, in addition to clinical risk factors such as radiation therapy, genetic variation can contribute to the POI risk. I considered both clinical and genetic components in the logistic regression. However, instead of estimating the effect for each clinical risk factor, a clinical risk score (CRS) is included in the model to represent the clinical risk(1). The CRS is introduced in the following paragraphs.

The CRS is the linear predictor -- a linear combination of the regression coefficients and the predictors -- in the previous risk prediction model for POI(1). To get a genuine estimate of the clinical risk, the effect of the CRS was adjusted by remodeling the relationship between the CRS and the outcome of interest on new data(4). The remodeling can be written as:

$$f(y_{new}) = a + b * linear\ predictor$$

The linear predictor is the CRS estimated from the previous clinical risk prediction model. The coefficient of the linear predictor (b) represents the amount of adjustment required to obtain a "true" effect of the clinical factors on POI. This remodeling process is called recalibration. and b is also known as the *calibration slope*(5). Ideally, b will have a value of 1, representing the perfect calibration. A calibration slope less than 1 indicates the amount we need to reduce on effects of predictors on average to make the model well-calibrated for new patients from the underlying population. In contrast, a calibration slope larger than 1 indicates the amount needed to increase for better calibration(4). The adjusted clinical risk score (CRS), i.e., b*linear predictor, is included in the logistic regression as an offset term. The PRS then is added to the model to account for the variations of POI risk that could not be explained by the clinical risk -- CRS.

We also considered ancestry as a confounding variable in the time-specific logistic regression, as ancestry is a determinant of the genetic structure and is often considered in genetic studies. The effect estimation of PRS may be spurious if the ancestry is not considered(6). The ancestry difference may come from the demographic history of a population, natural selection, and random fluctuations resulting from admixture(7). Principal components analysis (PCA) is often used to infer the population structure (ancestry difference)(8,9). The top principal components (PCs) obtained from PCA can be used as covariates in the modeling to account for the population structure(10). I included the first five PCs in the model to account for the ancestry.

The CRS was used as the baseline, written as:

$$log(p/(1-p)) = \text{CRS}$$

The main effect of PRSs and the interaction effect between PRSs and 1) the CRS; 2) radiation therapy (Yes/No) were examined. The conceptual models were written as:

Main effect model: $log(p/(1-p)) \sim$ PRS+ offset (CRS) +first five PCs

The interaction models: $log(p/(1-p)) \sim$ PRS*CRS + offset (CRS) +first five PCs

$log(p/(1-p)) \sim$ PRS*ovarian radiation therapy (RT: yes/no) + offset (CRS) + first five PCs

Where p represents the probability of developing POI. $p/(1-p)$ is the odds ratio, and the linear predictor in the right hand of the formula is proportional to $log(p/(1-p))$. PRSs calculated from both the general population and childhood cancer survivor GWA studies were examined in separate models.

A five-fold cross-validation framework was used for the internal validation(11). Model performance was assessed using the following metrics: the area under the receiver operating characteristic curve (AUC) was used

to measure the discrimination(12). The average positive predictive value (AP)(13) was used to measure the predictive accuracy. The Spiegelhalter-z statistic(14) was calculated to quantify the calibration. The overall performance was assessed using scaled Brier Score (sBrS). The 95% confidence intervals for AUC, AP, and sBrS were computed using bootstrapping resampling technique(12). Calibration curves were generated to visualize the performance by plotting the mean observed proportions with the event of each subgroup to its mean predicted probabilities. Finally, we compared the CRS model with the above-listed models and examined the incremental value of adding PRS in the prediction of POI. Analysis was performed using R version 4.0.3(15).

3.3 Results

Model Evaluation

The modeling and internal validation were based on the weighted samples (IPCW weights). A total of 2427 "participants" were used for the analysis after accounting for censoring weights, with 276 (11.40%) participants developing POI during the study period. The performance measurements, sBrS, AP, and AUC, plus Spiegelhalter-z statistic of the time-specific logistic regression models were computed. All the metrics were computed on the validation sets, and the results were then averaged over the validation sets.

The performance of all models is summarized in Table 3-1. The first row provides the performance of the CRS model as a baseline: the sBrS estimate was 0.236 (95% CI: 0.203-0.267); the AUC estimate was 0.797 (95% CI: 0.778-0.816), indicating adequate discrimination; and the AP value was considerably higher than the 0.114 event rate, reaching 0.539 (95% CI: 0.502-0.574). However, the Spiegelhalter-z statistic was 11.427 (95% CI: 9.467-13.387), indicating the calibration could be improved.

Values of all metrics were similar across different types of PRSs (i.e., *g*PRSs, *c*PRSs). The three *g*PRSs (PRS48, PRS69, and PRS262) performed similarly, so only the results for PRS69 are presented in this Chapter. The results for PRS48- and PRS262-based models are provided in Appendix J. Of twelve candidate *ct*PRSs, the *ct*PRS with *LD* R^2*=0.4 and P=5x10*$^{-5}$ performed best based on the average performance of AUC, AP and sBrS, and the Spiegelhalter-z statistic in the 5-fold internal cross-validation. Appendix K provided the results for all 12 candidate *ct*PRSs.) The *ct*PRS mentioned below is the *ct*PRS with *LD* R^2*=0.4 and P=5x10*$^{-5}$*,* which included nine genetic variants.

For the PRS main effect models, AUCs ranged from 0.775 to 0.780, and AP values ranged from 0.530 to 0.532. The AUC and AP values were similar to that of the CRS model, showing that none of the PRS main effect models improved discriminatory accuracy compared to the CRS model. The Spiegelhalter-z statistics ranged from 0.099 to 0.154, indicating the models were relatively well-calibrated. Compared to the CRS model, the Spiegelhalter-z statistics decreased from 11.427 to a range between 0.099 and 0.154, implying improved calibration in the PRS main effect models. The overall performance, measured by sBrS, improved with the improvement in calibration.

The AUC, AP, and sBrS estimates for the PRS*CRS models ranged from 0.792 to 0. 799, 0.528 to 0.537, and 0.227 to 0.257, respectively. These results were close to that of the CRS model, indicating that the PRS*CRS models did not improve discriminatory accuracy. Though the Spiegelhalter-z statistics, ranging from 3.774 to 6.686, decreased compared to the CRS model (Spiegelhalter-z statistic: 11.427), the 95% CIs, none of which included 0, suggested that the Spiegelhalter-z statistics were not statistically significant.

For the PRS*RT models, the AUCs were slightly lower than the CRS model (AUC: 0.797, 95%CI (0.778, 0.816)), ranging from 0.775 to 0.780. The AP values for the PRS*RT models ranged from 0.528 to 0.537, which remained similar to the CRS clinical model's AP value. So the PRS*RT models still did not improve the discriminatory accuracies compared to the CRS model. However, similar to the PRS main effect models, the Spiegelhalter-z statistics for the PRS*RT models decreased compared to the CRS model (a range of 1.401-1.485 vs. 11.427) and were statistically significant.

In summary, these metrics suggested that the PRS*CRS models did not improve the model performance. The PRS main effect models and PRS*RT models improved the overall performance of the models by improving the calibration (bolded in Table 3-1). However, none of the models improved the discrimination.

Table 3-1 Summary of model performance

			sBrS	AP	AUC	Spiegelhalter-z
CRS			0.236 (0.203, 0.267)	0.539 (0.502, 0.574)	0.797 (0.778, 0.816)	11.427 (9.467,13.387)
Main effect models		PRS69	**0.277 (0.242, 0.311)**	0.532 (0.495, 0.569)	0.780 (0.756, 0.804)	**0.134 (-1.826,2.094)**
		PRS6	**0.274 (0.240, 0.308)**	0.530 (0.494, 0.568)	0.775 (0.748, 0.801)	**0.154 (-1.806, 2.114)**
		*ct*PRS	**0.276 (0.242, 0.310)**	0.531 (0.495, 0.569)	0.776 (0.749, 0.801)	**0.099 (-1.861, 2.059)**
Interaction models	CRS with	PRS69	0.227 (0.180, 0.274)	0.537 (0.496, 0.578)	0.799 (0.780, 0.818)	6.686 (4.726, 8.646)
		PRS6	0.239 (0.193, 0.284)	0.534 (0.494, 0.575)	0.795 (0.776, 0.816)	5.781 (3.821, 7.741)
		*ct*PRS	0.257 (0.214, 0.299)	0.528 (0.486, 0.572)	0.792 (0.773, 0.813)	3.774 (1.814, 5.734)
	RT with	PRS69	**0.267 (0.229, 0.304)**	0.520 (0.481, 0.557)	0.752 (0.731, 0.774)	**1.485 (-0.475,3.445)**
		PRS6	**0.268 (0.230, 0.305)**	0.520 (0.481, 0.557)	0.752 (0.731, 0.774)	**1.401 (-0.559, 3.361)**
		*ct*PRS	**0.269 (0.231, 0.306)**	0.519 (0.478, 0.560)	0.755 (0.732, 0.779)	**1.142 (-0.818, 3.102)**

Note: Values are averaged over validation sets; ctPRS is the clumping and thresholding PRS, CRS represents the clinical risk score model, AUC is the area under the receiver operating characteristic curve, AP is the time-specific average positive predictive value. sBrS is the scaled Brier Score. Among all candidate ctPRSs constructed in Chapter 2, ctPRS with clumping = 0.4 and thresholding = 5e-5 was selected based on the averaged model performance. Detailed methodology and selection procedures were given in Appendix K. The metrics were accounted for IPCW weights

The Calibration Curve

The calibration of the models was assessed using Spiegelhalter-z statistic as aforementioned. Furthermore, calibration curves were plotted to visualize the results. Figure 3-1 and Figure 3-2 show the calibration curves. The x-axis shows the range of predicted probabilities of developing POI, and the y-axis reflects the observed proportions of POI in each subgroup.

Figure 3-1 shows the results for the CRS and PRS main effect models. The CRS model underestimated the POI risk, illustrated by the red line lying above the diagonal dashed line. The PRS69, PRS6, and *ct*PRS models performed similarly, with calibration curves close to the diagonal line. The PRS main effect models performed well for participants in medium risks but overestimated the risk for participants in high-risk groups (risk >0.5) and underestimated the risk for low-risk groups (risk <0.5). The PRS69 main effect model performed best in low-risk groups compared to the remaining models.

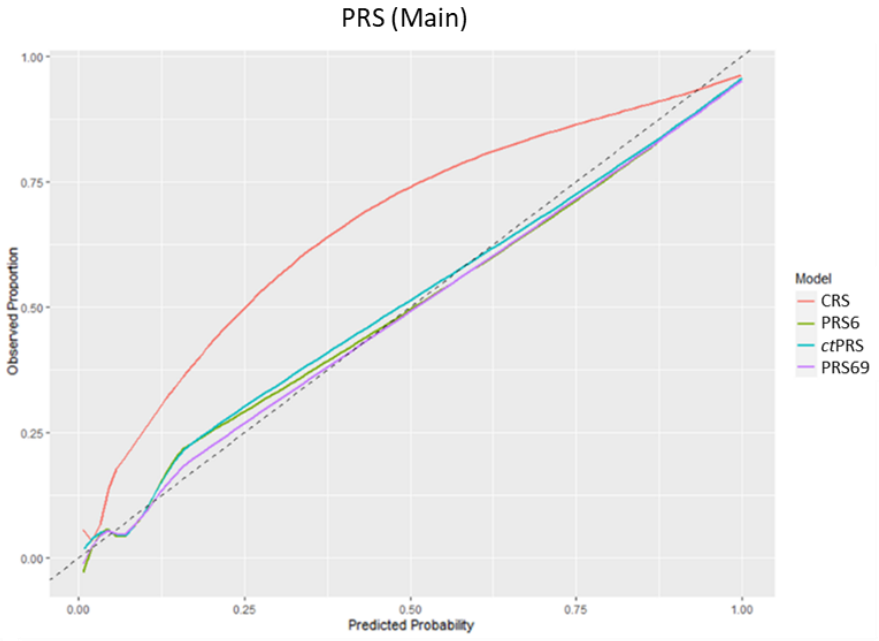
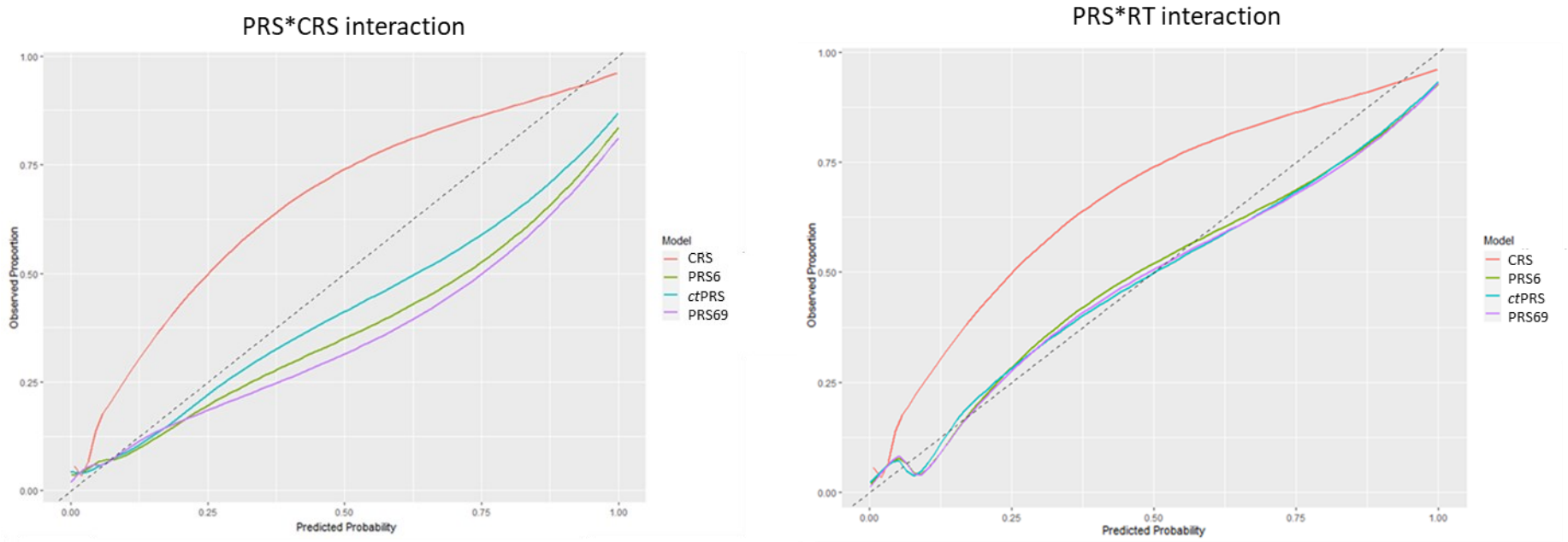


Figure 3-1 Calibration curves for the main effect models



Note: red line refers to the CRS clinical model, purple line refers to the PRS69 models, the green line refers to the PRS6 models, and blue line refers to the ctPRS models.

Figure 3-2 Calibration curves for the interaction models

Figure 3-2 shows the calibration curves for the interaction models. The PRS*CRS interaction models (Figure 3-2 left) performed well initially, but the curve was away from the diagonal line with the increase of the actual risk in participants, indicating overestimating the actual risks. Among the PRS*CRS models, the *ct*PRS*CRS model performed best, followed by the PRS6*CRS model, and finally, the PRS69*CRS model. The calibration curves of PRS*RT models (Figure 3-2 right) showed similar patterns with the PRS main effect models, with overestimated actual risk for participants in the high-risk groups and underestimated actual risk for participants in the low-risk groups.

Risk Stratification

Using 5%, 20%, and 50% as cutoffs, the predicted probabilities of developing POI were stratified into four risk categories, that is, <5%, 5% to <20%, 20% to <50%, and ⩾50%, corresponding to low, medium-low, medium, and high-risk groups, respectively.

Table 3-2 shows the risk stratification for the PRS main effect models. The risk stratification of the CRS showed that among 75 participants with a predicted risk greater than 50%, 84% developed POI. However, 22.4% of participants who developed POI were predicted as medium-low risk (5%-20%), indicating an underestimated risk for patients in this group. Among 69 participants who were classified into the medium-risk subgroup (20%-50%), 43 (63%) participants were diagnosed with POI.

The PPV value for the CRS model in the high-risk (>50%) subgroup was slightly higher than that of the rest models. The PPV value for the CRS model in the high-risk subgroup was 84%. The PRS69 main effect model performed best among PRS main effect models, with 107 cancer survivors having predicted risk greater than 0.5 (high-risk), among whom 82.24% of participants developed POI. The PPV values for the rest models were even smaller than that of the PRS69 main effect model.

Among participants who were predicted as medium risk (20%-50%) of developing POI in CRS, PRS69, PRS6, and *ct*PRS main effect models, 62%, 31%, 29%, and 29% of them developed POI, respectively. The interpretation of the results in the medium-risk group may differ regarding the different focuses in clinical practice. For example, if using a POI risk of 50% as the cutoff point for fertility preservation decision-making, The misclassification rate of the PRS main effect models in the subgroup is around 29%-31%. In contrast, the misclassification rate reached 62% for the CRS model. However, if a lower POI risk, say 20%, is of clinical significance, in this case, the CRS model can perform better in assisting the decision-making process.

The risk prediction for the medium low-risk (5%-20%) subgroup in the PRS main effect models was better compared to the CRS model (if we predict medium low-risk group as not going to develop POI). Among those 303 participants classified into the medium-low risk subgroup by the CRS model, 22% were misclassified. The PRS69 main effect model classified 1450 cancer survivors into the medium-low risk subgroup, among whom 93% did not develop POI (93% ≈ (1450-101)/1450), reducing the misclassification rate from 22% to 7%. The misclassification rates among the medium-low risk subgroup were even lower for the PRS6 and *ct*PRS main effect models (6%).

The risk prediction results were comparable among the models for the low-risk subgroup. The PPV value for the baseline model is 5.15%. And the PPV values ranged from 4.01% to 6.48% in the PRS main and interaction models.

In summary, the PRS main and interaction models generated similar results for risk stratification using the selected cutoffs. The risk stratification results in the low- and high-risk subgroups were similar among the CRS and PRS-based models. However, the PPV results for the medium-low and medium-risk subgroups differ between the CRS and PRS-based models.

Table 3-2 Risk stratification of the risk prediction models

Risk Categories			<5% (low-risk)			5%-20% (medium-low)			20%-50% (medium)			>50% (high-risk)		
			POI event	# of survivors	PPV(%)	POI event	# of survivors	PPV(%)	POI event	# of survivors	PPV(%)	POI event	# of survivors	PPV(%)
CRS			102	1980	5.15	68	303	22.44	43	69	62.32	63	75	84.00
Main Effect Models		PRS69	29	682	4.25	101	1450	6.97	58	188	30.85	88	107	82.24
		PRS6	36	629	5.72	99	1507	6.57	52	181	28.73	89	110	80.91
		*ct*PRS	37	662	5.59	96	1470	6.53	53	184	28.80	89	111	80.18
Interaction Models	CRS with	PRS69	65	1507	4.31	44	531	8.29	47	206	22.82	120	183	65.57
		PRS6	65	1468	4.43	49	607	8.07	50	184	27.17	112	168	66.67
		*ct*PRS	47	1273	3.69	65	815	7.98	55	189	29.10	109	150	72.67
	RT with	PRS69	36	898	4.01	95	1243	7.64	51	162	31.48	93	123	75.61
		PRS6	39	886	4.40	90	1247	7.22	54	174	31.03	92	119	77.31
		*ct*PRS	52	803	6.48	79	1328	5.95	53	181	29.28	91	115	79.13

Note: censoring weights (IPCW) were considered

The Coefficients and Significance

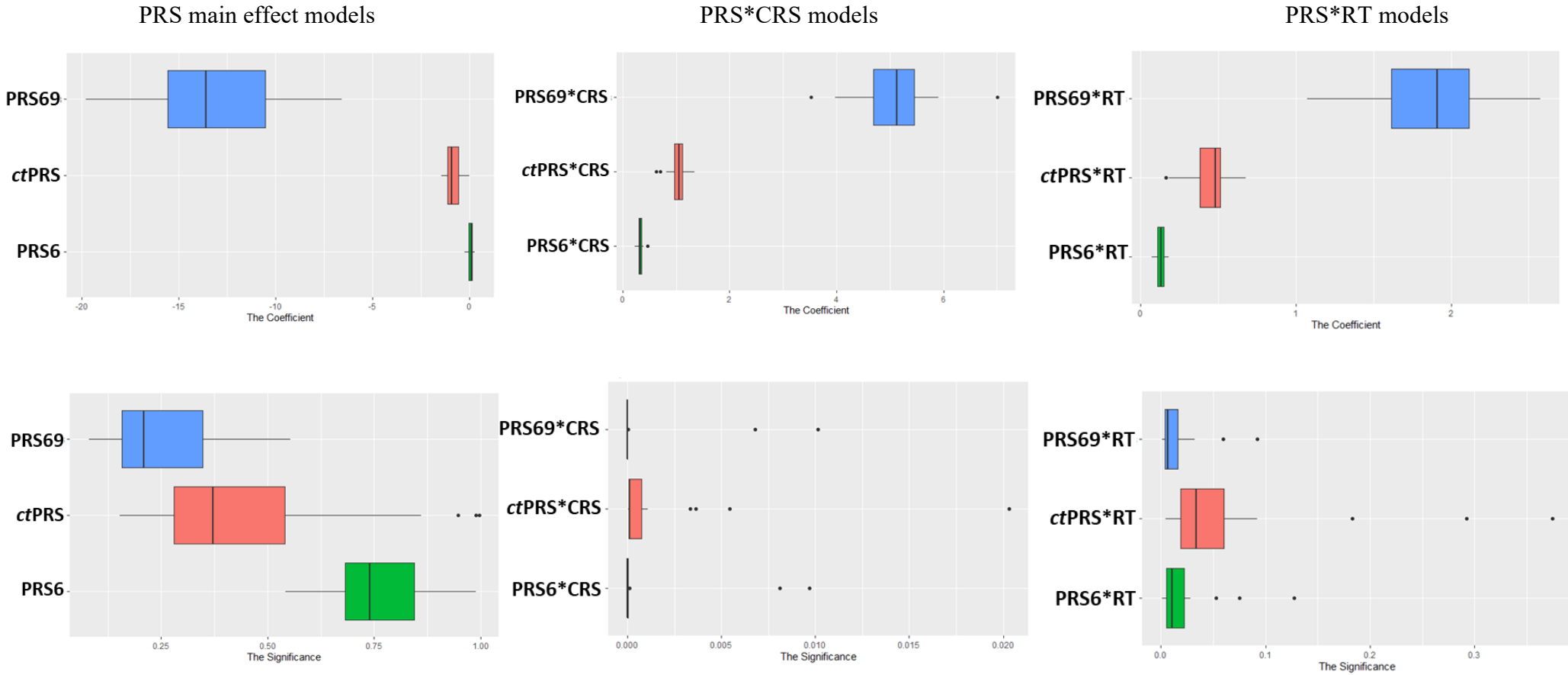
In addition to the predictive power, the significance of the genetic component is also of interest. We used a five-fold cross-validation framework (discussed in Section 1.2.3). We have five datasets generated from multiple imputation. Each dataset has five training sets, resulting in 25 training sets in total. Therefore, the modeling process was repeated 25 times. So instead of getting a single value for the parameters of interest (coefficient of PRS/PRS*CRS/PRS*RT), 25 estimates were generated for each parameter of interest. To present the point estimates, I summarized the means and medians for the coefficients and significance of PRS main effects and PRS*CRS/PRS*RT interaction effects in Table 3-3 and Table 3-4. I also provided the boxplots for the coefficients and significance in Figure 3-3 to show the variations.

Table 3-3 Summary of the coefficient estimates of PRSs in the training sets

	PRS69	**PRS6**	***ct*PRS**
PRS main			
Mean (SD)	-13.300 (3.420)	0.060 (0.154)	-0.785 (0.422)
Median [Min, Max]	-13.6 [-19.800, -6.590]	0.113 [-0.245, 0.281]	-0.904 [-1.430, -0.004]
PRS*CRS			
Mean (SD)	5.100 (0.688)	0.329 (0.0486)	1.030 (0.168)
Median [Min, Max]	5.13 [3.530, 7.010]	0.323 [0.228, 0.468]	1.05 [0.634, 1.340]
PRS*RT			
Mean (SD)	1.870 (0.359)	0.127 (0.027)	0.450 (0.126)
Median [Min, Max]	1.910 [1.070, 2.570]	0.131 [0.069, 0.180]	0.482 [0.163, 0.674]

Table 3-4 Summary of P-values associated with PRSs in the training sets

	PRS69	**PRS6**	***ct*PRS**
PRS main			
Mean (SD)	0.254 (0.136)	0.753 (0.123)	0.441 (0.270)
Median [Min, Max]	0.209 [0.0812, 0.553]	0.740 [0.541, 0.988]	0.371 [0.153, 0.997]
PRS*CRS			
Mean (SD)	0.681e-02 (0.00239)	0.727e-02 (0.00247)	0.15e-01(0.00415)
Median [Min, Max]	0.715 e-05 [0.308e-06, 0.0101]	0.322e-04 [0.105e-05, 0.00969]	0.123e-02 [0.179 e-03, 0.0203]
PRS*RT			
Mean (SD)	0.0150 (0.0205)	0.0206 (0.0280)	0.0633 (0.0900)
Median [Min, Max]	0.00692 [0.148e-02, 0.0924]	0.0105 [0.998e-02, 0.128]	0.0341 [0.484e-02, 0.374]



*Note: the first row is the results of the coefficients, the second row is the results of the significance; the first column displays the results for PRS main effect models, and the second column is the results for PRS*CRS models, the third column is the results for PRS*RT models.*

Figure 3-3 Boxplots of the coefficient estimates and their P-values of PRSs over the training sets

The second row of the boxplots in Figure 3-3 and Table 3-4 suggested that none of the PRSs in the main effect models were statistically significantly associated with POI after accounting for the CRS and the first five PCs. (P-value ranged from 0.25 to 0.75); all PRS*CRS interactions are statistically significant; all the PRS*RT interactions, except for the *ct*PRS*RT interaction, are also statistically significant.

For the significant associations (between PRS*CRS/PRS*RT and developing POI), The effect sizes for PRS69-based interactions were larger than that of the PRS6-based interactions. For example, on average, in the presence of ovarian radiation, a one-unit increase in the PRS69*RT will lead to a 6.5 times increase in the odds of developing POI (OR=exp(1.87)≈6.5, P-values <0.005). However, the PRS6*RT and *ct*PRS*RT models showed that the odds of developing POI among patients who received ovarian radiation therapy would increase by 13.54% and 56.83%, with one unit increase in the PRS6 and *ct*PRS, respectively. If a patient did not receive any ovarian radiation, the PRS would not affect the risk of developing POI. For the PRS*CRS interactions, a one-unit increase in the PRS69*CRS will lead to a 164-times increase in the odds of developing POI; and the PRS6 *CRS model showed that the odds of developing POI would increase by 38.96% with one unit increase in the PRS6*CRS on average. And one unit increase in the *ct*PRS*CRS will lead to a 2.8 times (2.8=exp (1.03)) increase in the odds of developing POI.

3.4 Discussion

This research investigated the added value of genetics in prediction models for POI by including genetic variants in the form of PRS. PRS generated from either the general population or cancer survivors were included in a time-specific logistic regression model. Both the main effect and interaction effect between PRS and treatments were examined.

The performance, from PRS main effect model to the PRS*CRS, and PRS*RT interaction models implied that after accounting for clinical risk factors, the inclusion of genetic data in the form of PRS could improve the overall performance of the predictive model for POI. Three different PRSs computed from GWAS conducted in the general population or childhood cancer survivors performed similarly. The predictive accuracy, which was captured by AP values, was similar across all models. The AUC values were also similar, with those in the CRS*PRS models being slightly lower than those in the other models. The Spiegelhalter-z statistics suggested that the addition of genetics improved the calibration.

One implication from this study is that the general population-based PRS which included 69 genetic variants, showed predictive power for POI in the childhood cancer survivor population. The calibration improved by having the PRS69 as the main effect in the model, though the main effect of PRS69 was not statistically significant in the model.

The study did not observe an association between POI and the PRS6 calculated from the six independent genetic variants reported from the GWA study of the childhood cancer survivor. In the previous GWAS study, though none of the genetic variants showed a statistically significant association with POI, the researchers suggested that the presence of a haplotype among patients exposed to ovarian RT may be at a high risk of developing POI. The PRS6*RT and *ct*PRS*RT interaction effects were found in this study, which further confirmed the conclusion in the previous study(16). However, the genetic variants included in the haplotype identified in the cancer survivor GWA study, PRS6, and *ct*PRS barely overlap.

The genetic information (i.e., genetic summary statistics) and clinical risk were obtained from the existing studies. The sample size of the GWAS for POI in childhood cancer survivors was limited, which may have resulted in the PRS not being representative of the genetic profile of POI. The GWAS study included only 799 cancer survivors with 30 (3.8%) cases. The limited sample size and the small number of cases might lead to effect estimates with large variance for the genetic variants. In the future, more extensive studies should be conducted to validate the PRS.

The *g*PRSs were computed from GWA studies in the general population and were applied to the cancer survivor population. Also, the *g*PRSs was derived from the GWA studies for age at natural menopause, which is not necessarily the same as POI. The validity of the *g*PRSs could be examined if some external datasets from the general population are available.

Another limitation was that half of the participants were censored in this study. The censoring weights were applied to the participants with observed outcomes to account for the censoring. For the application of the censoring weights, we assumed missing at random and independence between the event and censoring process. The censoring weights may not be valid if these assumptions are not met.

In conclusion, after controlling for clinical risk factors (represented by a CRS), the PRS-based genetic profile has shown prediction potential for POI in childhood cancer survivors. In comparison to the CRS clinical model, including general population-based PRS in the predictive model enhanced the calibration of the prediction models. Effect modification of PRS on ovarian radiation therapy were observed. External validation will be required in the future to confirm the findings. Ultimately, the predicted risk could be used for risk classification and as a quantitative reference for clinicians and patients when making fertility preservation decisions.

3.5 References

1. Lu Z. Risk Prediction for Premature Ovarian Insufficiency in Childhood Cancer Survivors. University of Alberta; 2020.

2. Khera A, Chaffin M, Aragam K, Haas M, Roselli C, Choi S, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nature Genetics. 2018 Sep 1;50.

3. Robison LL, Armstrong GT, Boice JD, Chow EJ, Davies SM, Donaldson SS, et al. The Childhood Cancer Survivor Study: A National Cancer Institute–Supported Resource for Outcome and Intervention Research. J Clin Oncol. 2009 May 10;27(14):2308–18.

4. Steyerberg EW. Clinical prediction models. Springer; 2019.

5. Boqué R, Rius FX, Massart DL. Straight line calibration: something more than slopes, intercepts, and correlation coefficients. Journal of chemical education. 1994;71(3):230.

6. Zhu X, Tang H, Risch N. Admixture mapping and the role of population structure for localizing disease genes. Adv Genet. 2008;60:547–69.

7. Jin J, Cerise JE, Kang SJ, Yoon EJ, Yoon S, Mendell NR, et al. Principal components ancestry adjustment for Genetic Analysis Workshop 17 data. BMC Proc. 2011 Nov 29;5(Suppl 9):S66.

8. Menozzi P, Piazza A, Cavalli-Sforza L. Synthetic Maps of Human Gene Frequencies in Europeans. Science. 1978;201(4358):786–92.

9. Ma J, Amos C. Theoretical Formulation of Principal Components Analysis to Detect and Correct for Population Stratification. PloS one. 2010 Sep 17;5.

10. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006 Aug;38(8):904–9.

11. Refaeilzadeh P, Tang L, Liu H. Cross-validation. Encyclopedia of database systems. 2009;5:532–8.

12. Huang J, Ling CX. Using AUC and accuracy in evaluating learning algorithms. IEEE Transactions on knowledge and Data Engineering. 2005;17(3):299–310.

13. Yuan Y, Su W, Zhu M. Threshold-Free Measures for Assessing the Performance of Medical Screening Tests. Frontiers in Public Health. 2015;3:57.

14. Spiegelhalter DJ. Probabilistic prediction in patient management and clinical trials. Statistics in Medicine. 1986;5(5):421–33.

15. Ihaka R, Gentleman R. R: a language for data analysis and graphics. Journal of computational and graphical statistics. 1996;5(3):299–314.

16. Brooke RJ, Chemaitilly W, Wilson CL, Krasin MJ, Li Z, Im C, et al. A high-risk genetic profile for premature menopause (PM) in childhood cancer survivors (CCS) exposed to gonadotoxic therapy: A report from the St. Jude Lifetime Cohort (SJLIFE) and Childhood Cancer Survivor Study (CCSS). JCO. 2017 May 20;35(15):10502–10502.

17. Green DM, Sklar CA, Boice Jr JD, Mulvihill JJ, Whitton JA, Stovall M, et al. Ovarian failure and reproductive outcomes after childhood cancer treatment: results from the Childhood Cancer Survivor Study. Journal of Clinical Oncology. 2009;27(14):2374.

18. Mahmoodi N, Bekker H, King N, Hughes J, Jones G. Decision Aids' Efficacy to Support Women's Fertility Preservation Choices Before Cancer Treatment: An Environmental Scan. In: ISDM 2017 Abstract Book: Oral communications. Leeds; 2017.

19. Bzdok D, Engemann D, Thirion B. Inference and Prediction Diverge in Biomedicine. Patterns (N Y). 2020 Oct 8;1(8):100119.

20. Rossetti R, Ferrari I, Bonomi M, Persani L. Genetics of primary ovarian insufficiency. Clin Genet. 2017 Feb;91(2):183–98.

21. Dudbridge F. Power and predictive accuracy of polygenic risk scores. PLoS Genet. 2013 Mar;9(3):e1003348.

22. Choi SW, Mak TS-H, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. Nat Protoc. 2020 Sep;15(9):2759–72.

23. Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M. Genome-wide association studies in diverse populations. Nature Reviews Genetics. 2010;11(5):356–66.

24. Hickey PF, Bahlo M. X chromosome association testing in genome wide association studies. Genetic epidemiology. 2011;35(7):664–70.

25. Coughlin SS. Recall bias in epidemiologic studies. Journal of clinical epidemiology. 1990;43(1):87–91.

26. Leung K-M, Elashoff RM, Afifi AA. Censoring issues in survival analysis. Annual review of public health. 1997;18(1):83–104.

27. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests. Biometrics. 2000 Sep;56(3):779–88.

28. Wallace WHB, Smith AG, Kelsey TW, Edgar AE, Anderson RA. Fertility preservation for girls and young women with cancer: population-based validation of criteria for ovarian tissue cryopreservation. Lancet Oncol. 2014 Sep;15(10):1129–36.

4 Conclusions

4.1 Summary

Primary ovarian insufficiency is a major concern in female childhood cancer survivors(1). The POI risk prediction models would help clinicians identify individuals at elevated risk of the condition, thus informing clinical decision-making regarding fertility preservation and improving the long-term quality of life for childhood cancer survivors(2). The treatment-related risks have been developed for POI and showed potential for risk stratification(3). This study aims to investigate whether the model performance can be improved by incorporating the genetic profile on the basis of the existing clinical risk model.

In Chapter 1, I reviewed researches of clinical risk factors, genetic architecture, and the clinical risk prediction model for POI. I also introduced the construction of the PRS, and the modeling and evaluation of risk prediction models. Finally, the CCSS original cohort was introduced as the main data source of this study.

In Chapter 2, GWA studies were extracted, and sixteen PRSs were constructed: three *g*PRSs (PRS48, PRS69 and PRS262) were built from two GWA studies for age at natural menopause in the general population; a GWA study for POI conducted in the SJLIFE cohort was used to construct thirteen *c*PRSs (PRS6 and twelve *ct*PRSs), where PRS6 was constructed from the top 6 independent genetic variants (with $P<10^{-5}$) and *ct*PRSs were constructed using the C+T method.

In Chapter 3, PRS main effect models were examined using the time-specific logistic regression at age 40 after accounting for the CRS and the first five PCs. These include PRS69, PRS6, and *ct*PRS main effect models. Potential interactions including PRS* CRS and PRS*RT were also examined. A five-fold cross-validation framework was applied for the model building and performance evaluation. The time-specific logistic regression models were developed on the training set. The AP, AUC, sBrS values, calibration curves, and the

Spiegelhalter-z Statistic were calculated on the validation datasets to assess the model performance. The coefficient estimates and associated P-values for these PRSs were also presented in Chapter 3.

The model performance of the PRS main effect models was similar. The PRS main effect models performed similarly to the CRS model regarding the AUC and AP values. The discrimination did not improve compared to the CRS model. However, the calibration curves and the Spiegelhalter-z Statistics suggested improvements in the calibration in the PRS main effect models. The PRS69 main effect model was able to identify more POI cases in survivors predicted to be high POI risk. The PRS*CRS interaction models overestimated the risk in general, and the PRS*RT, especially the *ct*PRS*RT model performed well.

4.2 Study Limitations

There is a lack of external data from the general population for the development of general population-based PRS. As a result, only significant independent genetic variations reported in GWA studies were used to create the general population-based PRSs, meaning that information of the remaining genetic variants which might have some predictive power for POI was not considered. Moreover, the statistical significance does not necessarily guarantee a higher predictive power(4). Genetic variants which are less significant could be considered if external data is available in the future. Also, the general population-based PRSs were developed from GWA studies for age at natural menopause instead of POI. The idea of applying PRS for age at menopause on the POI risk prediction model was based on the assumption that POI and age at natural menopause shared a similar genetic architecture(5). Though the assumption has been expressed by researchers and was reasonable, GWA studies for POI in the general population could be used if available.

The effect size estimates of the genetic variants in the GWA study conducted in the female childhood cancer survivors might not be accurate due to the limited sample size(6). Thus, the improvement observed in the

calibration should be interpreted with caution. Moving forward, the developed model needed to be applied to external data to assess its validity. Moreover, the GWA study for POI in the cancer survivor population could be conducted when more GWAS data is available. Consequently, the PRS for POI in the population of female childhood cancer survivors could be reconstructed with more confidence.

The effect estimate of the *ct*PRS in the *ct*PRS main effect model showed the opposite direction to effect estimates of the genetic variants in the GWAS study. Theoretically, the effect estimates of both the *ct*PRSs and the genetic variants should be in the same direction, as the PRS is simply a sum of effect sizes of all the variants(7). The difference may come from the following aspects. First, both analyses have accounted for the clinical component to study the genetics of POI in childhood cancer survivors. However, the clinical component was considered differently between the original GWAS study and my study. The clinical component in my study entered the model as a clinical risk score which was modeled from the demographic information, chemotherapy agents, and radiation dosage to ovary information, whereas the GWAS study utilized the cyclophosphamide equivalent dose of alkylating agents (CED) and ovarian radiotherapy (Yes/No) to account for the clinical risk(8). Second, the GWA study used a continuous variable to account for the ancestry, but my study used the first five PCs. Moreover, the summary statistics — specifically, effect sizes and P-values—differ between the published GWAS study and the received summary statistics data(8). The GWA study reported the likelihood ratio test for the association between genetic variants and POI, but the received summary statistics data contained the Wald test, which reported similar effect estimates as the likelihood ratio test but different P-values (the significance of the genetic variants was different when the sample size was relatively small, which is the case in the original GWA study) for each genetic variant. The same problem was also present for the *c*PRS. The results in Chapter 3 suggested that the increase of the PRS could possibly decrease the risk of POI, though the effect was not statistically significant.

This study focused on European ancestry, thus the PRS and risk prediction models, which inform interventions, are more likely to benefit people of European ancestry (which account for 16% of the total global population). The PRS could be inaccurate when being applied to other populations. Currently, around 80% of GWA studies are conducted on people of European descent, resulting in a significant bias in the population that benefits(9). The bias can be corrected only by conducting further GWAS research in different racial groups, and more people will benefit from the PRS.

The genetic variants on chromosome X were not included in the PRS construction for POI, despite the fact that chromosome X variations explained some of the variances for age at natural menopause/POI(5). Because of the uniqueness of chromosome X, most GWA studies ignore sex chromosomes and focus solely on autosomal genetic variations in their analyses(10), which is the case for menopause-related phenotypes studies. Therefore, the exclusion of genetic variations that happened on chromosome X might have limited the potential of the PRS.

The age at the last menstrual period was self-reported, and recall bias could affect the time at risk for POI, thus affecting the modeling results(11). Besides, about half of the study participants in this study were censored. The time period for childhood cancer survivors to remain at risk of POI before age 40 can be up to 40 years. Thus the probability of censoring is high(12). The censored individuals did not contribute directly to the estimation of risk but contributed information to estimate the IPCW weights(13). Individuals with POI status assessed at an older age were given higher weights to account for the fact that they are more likely to be censored before experiencing menopause. The validity of the IPCW weights is based on the assumption that the censoring is independent of the event time given covariates(13).

4.3 Recommendations for Future Directions and Applications

Though the PRSs improved the calibration, external validation is necessary to confirm the validity of the risk predictive models before applying them in clinical practice. Following the external validation, the predicted risks could be categorized into different risk levels to inform the decision-making process.

The genetic variants included in the *g*PRSs and *c*PRSs do not overlap. This difference may be due to the different genetic architectures in these two populations or the limited GWA studies in childhood cancer survivors. The comparison between the *g*PRSs and *c*PRSs would be more accessible when *c*PRSs were being constructed with larger GWA studies in the childhood cancer survivors.

The effect modification of RT on the PRS was observed, showing a possible gene-treatment interaction. Collaboration with doctors is required to understand the effect modification of genetics on RT. Moreover, the interaction between PRS/single gene and other clinical risk factors (such as chemotherapy dosage and diagnosis type) is worth investigating in collaboration with clinicians.

4.4 References

1. Green DM, Sklar CA, Boice Jr JD, Mulvihill JJ, Whitton JA, Stovall M, et al. Ovarian failure and reproductive outcomes after childhood cancer treatment: results from the Childhood Cancer Survivor Study. Journal of Clinical Oncology. 2009;27(14):2374.

2. Mahmoodi N, Bekker H, King N, Hughes J, Jones G. Decision Aids' Efficacy to Support Women's Fertility Preservation Choices Before Cancer Treatment: An Environmental Scan. In: ISDM 2017 Abstract Book: Oral communications. Leeds; 2017.

3. Lu Z. Risk Prediction for Premature Ovarian Insufficiency in Childhood Cancer Survivors. University of Alberta; 2020.

4. Bzdok D, Engemann D, Thirion B. Inference and Prediction Diverge in Biomedicine. Patterns (N Y). 2020 Oct 8;1(8):100119.

5. Rossetti R, Ferrari I, Bonomi M, Persani L. Genetics of primary ovarian insufficiency. Clin Genet. 2017 Feb;91(2):183–98.

6. Dudbridge F. Power and predictive accuracy of polygenic risk scores. PLoS Genet. 2013 Mar;9(3):e1003348.

7. Choi SW, Mak TS-H, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. Nat Protoc. 2020 Sep;15(9):2759–72.

8. Brooke RJ, Chemaitilly W, Wilson CL, Krasin MJ, Li Z, Im C, et al. A high-risk genetic profile for premature menopause (PM) in childhood cancer survivors (CCS) exposed to gonadotoxic therapy: A report from the St. Jude Lifetime Cohort (SJLIFE) and Childhood Cancer Survivor Study (CCSS). JCO. 2017 May 20;35(15):10502–10502.

9. Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M. Genome-wide association studies in diverse populations. Nature Reviews Genetics. 2010;11(5):356–66.

10. Hickey PF, Bahlo M. X chromosome association testing in genome wide association studies. Genetic epidemiology. 2011;35(7):664–70.

11. Coughlin SS. Recall bias in epidemiologic studies. Journal of clinical epidemiology. 1990;43(1):87–91.

12. Leung K-M, Elashoff RM, Afifi AA. Censoring issues in survival analysis. Annual review of public health. 1997;18(1):83–104.

13. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests. Biometrics. 2000 Sep;56(3):779–88.

14. Wallace WHB, Smith AG, Kelsey TW, Edgar AE, Anderson RA. Fertility preservation for girls and young women with cancer: population-based validation of criteria for ovarian tissue cryopreservation. Lancet Oncol. 2014 Sep;15(10):1129–36.

References

Adriaens, I., Smitz, J., & Jacquet, P. (2009). The current knowledge on radiosensitivity of ovarian follicle development stages. Human Reproduction Update, 15(3), 359–377. https://doi.org/10.1093/humupd/dmn063

Baker, V. L. (2013). Primary ovarian insufficiency in the adolescent. Current Opinion in Obstetrics and Gynecology, 25(5), 375–381. https://doi.org/10.1097/GCO.0b013e328364ed2a

Biospecimens | St. Jude CCSS. (n.d.). Retrieved October 16, 2021, from https://ccss.stjude.org/biospecimens.html

Boqué, R., Rius, F. X., & Massart, D. L. (1994). Straight line calibration: Something more than slopes, intercepts, and correlation coefficients. Journal of Chemical Education, 71(3), 230.

Bosch, E., Alviggi, C., Lispi, M., Conforti, A., Hanyaloglu, A. C., Chuderland, D., Simoni, M., Raine-Fenning, N., Crépieux, P., Kol, S., Rochira, V., D'Hooghe, T., & Humaidan, P. (2021). Reduced FSH and LH action: Implications for medically assisted reproduction. Human Reproduction, 36(6), 1469–1480. https://doi.org/10.1093/humrep/deab065

Brooke, R. J., Chemaitilly, W., Wilson, C. L., Krasin, M. J., Li, Z., Im, C., Morton, L. M., Wu, G., Wang, Z., Chen, W., Howell, R. M., Armstrong, G. T., Bhatia, S., Chanock, S. J., Zhang, J., Green, D. M., Sklar, C. A., Hudson, M. M., Robison, L. L., & Yasui, Y. (2017). A high-risk genetic profile for premature menopause (PM) in childhood cancer survivors (CCS) exposed to gonadotoxic therapy: A report from the St. Jude Lifetime Cohort (SJLIFE) and Childhood Cancer Survivor Study (CCSS). JCO, 35(15), 10502–10502. https://doi.org/10.1200/JCO.2017.35.15_suppl.10502

Buccal Cell Collection. (n.d.). St. Jude CCSS. Retrieved October 31, 2021, from https://ccss.stjude.org/biospecimens/buccal-cell-collection.html

Bzdok, D., Engemann, D., & Thirion, B. (2020). Inference and Prediction Diverge in Biomedicine. Patterns, 1(8), 100119. https://doi.org/10.1016/j.patter.2020.100119

Cabry, R., Merviel, P., Hazout, A., Belloc, S., Dalleac, A., Copin, H., & Benkhalifa, M. (2014). Management of infertility in women over 40. Maturitas, 78(1), 17–21. https://doi.org/10.1016/j.maturitas.2014.02.014

Canada, P. H. A. of. (2012, July 9). Cancer in Children in Canada (0-14 years) [Research]. https://www.canada.ca/en/public-health/services/chronic-diseases/cancer/cancer-children-canada-0-14-years.html

Chemaitilly, W., Li, Z., Krasin, M. J., Brooke, R. J., Wilson, C. L., Green, D. M., Klosky, J. L., Barnes, N., Clark, K. L., & Farr, J. B. (2017). Premature ovarian insufficiency in childhood cancer survivors: A report from the St. Jude Lifetime Cohort. The Journal of Clinical Endocrinology & Metabolism, 102(7), 2242–2250.

Chian, R.-C., & Quinn, P. (2010). Fertility Cryopreservation. Cambridge University Press.

Childhood Cancer Survivor Study—2013 Progress Report. (2013). Childhood Cancer Survivor Study. https://ccss.stjude.org/content/dam/en_US/shared/ccss/documents/progress-report-2013.pdf

Choi, I. Y., Choi, Y. E., Nam, H. R., Lee, J. W., Park, E. C., & Jang, S. I. (2018). Relationship between Early Menopause and Mental Health Problems. Korean Journal of Family Practice, 8(1), 87–92. https://doi.org/10.21215/kjfp.2018.8.1.87

Choi, S. W., Mak, T. S.-H., & O'Reilly, P. F. (2020). Tutorial: A guide to performing polygenic risk score analyses. Nature Protocols, 15(9), 2759–2772. https://doi.org/10.1038/s41596-020-0353-1

Christin-Maitre, S., & Tachdjian, G. (2010). Genome-wide association study and premature ovarian failure. Annales d Endocrinologie, 71(3), 218–221.

Clark, R. A., Mostoufi-Moab, S., Yasui, Y., Vu, N. K., Sklar, C. A., Motan, T., Brooke, R. J., Gibson, T. M., Oeffinger, K. C., Howell, R. M., Smith, S. A., Lu, Z., Robison, L. L., Chemaitilly, W., Hudson, M. M., Armstrong, G. T., Nathan, P. C., & Yuan, Y. (2020). Predicting acute ovarian failure in female survivors of childhood cancer: A cohort study in the Childhood Cancer Survivor Study (CCSS) and the St Jude Lifetime Cohort (SJLIFE). The Lancet Oncology, 21(3), 436–445. https://doi.org/10.1016/S1470-2045(19)30818-6

Collins, F. S., & Varmus, H. (2015). A new initiative on precision medicine. New England Journal of Medicine, 372(9), 793–795.

Coughlin, S. S. (1990). Recall bias in epidemiologic studies. Journal of Clinical Epidemiology, 43(1), 87–91.

Data changes that occur between builds. (2005). National Center for Biotechnology Information (US). https://www.ncbi.nlm.nih.gov/books/NBK44467/

Day, F. R., Ruth, K. S., Thompson, D. J., Lunetta, K. L., Pervjakova, N., Chasman, D. I., Stolk, L., Finucane, H. K., Sulem, P., Bulik-Sullivan, B., Esko, T., Johnson, A. D., Elks, C. E., Franceschini, N., He, C., Altmaier, E., Brody, J. A., Franke, L. L., Huffman, J. E., … Murray, A. (2015). Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. Nature Genetics, 47(11), 1294–1303. https://doi.org/10.1038/ng.3412

Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. PLoS Genetics, 9(3), e1003348. https://doi.org/10.1371/journal.pgen.1003348

Dudbridge, F. (2016). Polygenic Epidemiology. Genetic Epidemiology, 40(4), 268–272. https://doi.org/10.1002/gepi.21966

Ensembl genome browser 104. (n.d.). Retrieved August 13, 2021, from https://uswest.ensembl.org/index.html

Faubion, S. S., Kuhle, C. L., Shuster, L. T., & Rocca, W. A. (2015). Long-term health consequences of premature or early menopause and considerations for management. Climacteric, 18(4), 483–491. https://doi.org/10.3109/13697137.2015.1020484

Gelson, E., Prakash, A., Macdougall, J., & Williams, D. (2016). Reproductive health in female survivors of childhood cancer. The Obstetrician & Gynaecologist, 18(4), 315–322. https://doi.org/10.1111/tog.12338

Genome browser FAQ. (n.d.). Retrieved August 13, 2021, from https://genome.ucsc.edu/FAQ/FAQreleases.html#release1

Genome wide association study ATLAS. (n.d.). Retrieved October 28, 2021, from https://atlas.ctglab.nl/traitDB/3366

Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations | Nature Genetics. (n.d.). Retrieved August 10, 2021, from https://www.nature.com/articles/s41588-018-0183-z

Glahn, H. R., & Jorgensen, D. L. (1970). CLIMATOLOGICAL ASPECTS OF THE BRIER P-SCORE. Monthly Weather Review, 98(2), 136–141. https://doi.org/10.1175/1520-0493(1970)098<0136:CAOTBP>2.3.CO;2

Gold, E. B., Bromberger, J., Crawford, S., Samuels, S., Greendale, G. A., Harlow, S. D., & Skurnick, J. (2001). Factors associated with age at natural menopause in a multiethnic sample of midlife women. American Journal of Epidemiology, 153(9), 865–874. https://doi.org/10.1093/aje/153.9.865

Gonçalves, V., Sehovic, I., & Quinn, G. (2014). Childbearing attitudes and decisions of young breast cancer survivors: A systematic review. Human Reproduction Update, 20(2), 279–292. https://doi.org/10.1093/humupd/dmt039

Green, D. M., Sklar, C. A., Boice Jr, J. D., Mulvihill, J. J., Whitton, J. A., Stovall, M., & Yasui, Y. (2009). Ovarian failure and reproductive outcomes after childhood cancer treatment: Results from the Childhood Cancer Survivor Study. Journal of Clinical Oncology, 27(14), 2374.

GWAS catalog. (n.d.). Retrieved August 11, 2021, from https://www.ebi.ac.uk/gwas/

Haller-Kikkatalo, K., Uibo, R., Kurg, A., & Salumets, A. (2015). The prevalence and phenotypic characteristics of spontaneous premature ovarian failure: A general population registry-based study. Human Reproduction, 30(5), 1229–1238.

He, C., Kraft, P., Chen, C., Buring, J. E., Paré, G., Hankinson, S. E., Chanock, S. J., Ridker, P. M., Hunter, D. J., & Chasman, D. I. (2009). Genome-wide association studies identify loci associated with age at menarche and age at natural menopause. Nature Genetics, 41(6), 724–728. https://doi.org/10.1038/ng.385

Hickey, P. F., & Bahlo, M. (2011). X chromosome association testing in genome wide association studies. Genetic Epidemiology, 35(7), 664–670.

Hu, B., Palta, M., & Shao, J. (2006). Properties of R2 statistics for logistic regression. Statistics in Medicine, 25(8), 1383–1395. https://doi.org/10.1002/sim.2300

Huang, J., & Ling, C. X. (2005). Using AUC and accuracy in evaluating learning algorithms. IEEE Transactions on Knowledge and Data Engineering, 17(3), 299–310.

Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics. Journal of Computational and Graphical Statistics, 5(3), 299–314.

Index of /goldenPath/hg18/database. (n.d.). Retrieved August 11, 2021, from http://hgdownload.soe.ucsc.edu/goldenPath/hg18/database/

Jin, J., Cerise, J. E., Kang, S. J., Yoon, E. J., Yoon, S., Mendell, N. R., & Finch, S. J. (2011). Principal components ancestry adjustment for Genetic Analysis Workshop 17 data. BMC Proceedings, 5(Suppl 9), S66. https://doi.org/10.1186/1753-6561-5-S9-S66

Jones, G., Hughes, J., Mahmoodi, N., Smith, E., Skull, J., & Ledger, W. (2017). What factors hinder the decision-making process for women with cancer and contemplating fertility preservation treatment? Human Reproduction Update, 23(4), 433–457. https://doi.org/10.1093/humupd/dmx009

Khera, A., Chaffin, M., Aragam, K., Haas, M., Roselli, C., Choi, S., Natarajan, P., Lander, E., Lubitz, S., Ellinor, P., & Kathiresan, S. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nature Genetics, 50. https://doi.org/10.1038/s41588-018-0183-z

Khera, A. V., Emdin, C. A., Drake, I., Natarajan, P., Bick, A. G., Cook, N. R., Chasman, D. I., Baber, U., Mehran, R., & Rader, D. J. (2016). Genetic risk, adherence to a healthy lifestyle, and coronary disease. New England Journal of Medicine, 375(24), 2349–2358.

Kirkman, M., Winship, I., Stern, C., Neil, S., Mann, G. B., & Fisher, J. R. W. (2014). Women's reflections on fertility and motherhood after breast cancer and its treatment. European Journal of Cancer Care, 23(4), 502–513. https://doi.org/10.1111/ecc.12163

Konuma, T., & Okada, Y. (2021). Statistical genetics and polygenic risk score for precision medicine. Inflammation and Regeneration, 41(1), 18. https://doi.org/10.1186/s41232-021-00172-9

Lambert, S. A., Gil, L., Jupp, S., Ritchie, S. C., Xu, Y., Buniello, A., McMahon, A., Abraham, G., Chapman, M., Parkinson, H., Danesh, J., MacArthur, J. A. L., & Inouye, M. (2021). The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. Nature Genetics, 53(4), 420–425. https://doi.org/10.1038/s41588-021-00783-5

Laven, J. S. E. (2015). Genetics of Early and Normal Menopause. Seminars in Reproductive Medicine, 33(6), 377–383.

Lawson, A. K., Klock, S. C., Pavone, M. E., Hirshfeld-Cytron, J., Smith, K. N., & Kazer, R. R. (2014). Prospective study of depression and anxiety in female fertility preservation and infertility patients. Fertility and Sterility, 102(5), 1377–1384. https://doi.org/10.1016/j.fertnstert.2014.07.765

Leung, K.-M., Elashoff, R. M., & Afifi, A. A. (1997). Censoring issues in survival analysis. Annual Review of Public Health, 18(1), 83–104.

Levine, J. M., Whitton, J. A., Ginsberg, J. P., Green, D. M., Leisenring, W. M., Stovall, M., Robison, L. L., Armstrong, G. T., & Sklar, C. A. (2018). Nonsurgical premature menopause and reproductive implications in survivors of childhood cancer: A report from the Childhood Cancer Survivor Study: NSPM and Reproductive Implications. Cancer, 124(5), 1044–1052. https://doi.org/10.1002/cncr.31121

Lewis, C. M., & Vassos, E. (2020). Polygenic risk scores: From research tools to clinical instruments. Genome Medicine, 12(1), 44. https://doi.org/10.1186/s13073-020-00742-5

Lift genome annotations. (n.d.). Retrieved August 13, 2021, from http://genome.ucsc.edu/cgi-bin/hgLiftOver

Louwers, Y. V., & Visser, J. A. (2021). Shared Genetics Between Age at Menopause, Early Menopause, POI and Other Traits. Frontiers in Genetics, 12, 1889. https://doi.org/10.3389/fgene.2021.676546

Lu, Z. (2020). Risk Prediction for Premature Ovarian Insufficiency in Childhood Cancer Survivors. University of Alberta.

Lund, L., Schmiegelow, K., Rechnitzer, C., & Johansen, C. (2011). A Systematic Review of Studies on Psychosocial Late Effects of Childhood Cancer: Structures of Society and Methodological Pitfalls May Challenge the Conclusions. Pediatric Blood & Cancer, 56, 532–543. https://doi.org/10.1002/pbc.22883

Ma, J., & Amos, C. (2010). Theoretical Formulation of Principal Components Analysis to Detect and Correct for Population Stratification. PloS One, 5. https://doi.org/10.1371/journal.pone.0012510

Mahmoodi, N., Bekker, H., King, N., Hughes, J., & Jones, G. (2017). Decision Aids' Efficacy to Support Women's Fertility Preservation Choices Before Cancer Treatment: An Environmental Scan. ISDM 2017 Abstract Book: Oral Communications.

Marwick, C. (2003). Childhood cancer survivors experience long term side effects. BMJ : British Medical Journal, 327(7414), 522.

Menozzi, P., Piazza, A., & Cavalli-Sforza, L. (1978). Synthetic Maps of Human Gene Frequencies in Europeans. Science, 201(4358), 786–792.

Miro, F., Parker, S. W., Aspinall, L. J., Coley, J., Perry, P. W., & Ellis, J. E. (2005). Sequential classification of endocrine stages during reproductive aging in women: The FREEDOM study. Menopause-the Journal of The North American Menopause Society, 12(3), 281–290. https://doi.org/10.1097/01.gme.0000147018.30796.25

Monthly Weather Review. (1950). War Department, Office of the Chief Signal Officer.

Moons, K. G. M., Altman, D. G., Reitsma, J. B., Ioannidis, J. P. A., Macaskill, P., Steyerberg, E. W., Vickers, A. J., Ransohoff, D. F., & Collins, G. S. (2015). Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): Explanation and elaboration. Annals of Internal Medicine, 162(1), W1-73. https://doi.org/10.7326/M14-0698

Mullins, N., Power, R. A., Fisher, H. L., Hanscombe, K. B., Euesden, J., Iniesta, R., Levinson, D. F., Weissman, M. M., Potash, J. B., & Shi, J. (2016). Polygenic interactions with environmental adversity in the aetiology of major depressive disorder. Psychological Medicine, 46(4), 759–770.

Nelson, L. M. (2009). Primary Ovarian Insufficiency. The New England Journal of Medicine, 360(6), 606–614. https://doi.org/10.1056/NEJMcp0808697

Perry, J. R. B., Corre, T., Esko, T., Chasman, D. I., Fischer, K., Franceschini, N., He, C., Kutalik, Z., Mangino, M., Rose, L. M., Vernon Smith, A., Stolk, L., Sulem, P., Weedon, M. N., Zhuang, W. V., Arnold, A., Ashworth, A., Bergmann, S., Buring, J. E., … Murray, A. (2013). A genome-wide association study of early menopause and the combined impact of identified variants. Human Molecular Genetics, 22(7), 1465–1472.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics, 38(8), 904–909. https://doi.org/10.1038/ng1847

Privé, F., Vilhjálmsson, B. J., Aschard, H., & Blum, M. G. B. (2019). Making the Most of Clumping and Thresholding for Polygenic Scores. The American Journal of Human Genetics, 105(6), 1213–1221. https://doi.org/10.1016/j.ajhg.2019.11.001

Pu, D., Xing, Y., Gao, Y., Gu, L., & Wu, J. (2014). Gene variation and premature ovarian failure: A meta-analysis. European Journal of Obstetrics, Gynecology, and Reproductive Biology, 182, 226–237. https://doi.org/10.1016/j.ejogrb.2014.09.036

Public access GWAS data tables. (n.d.). Retrieved August 7, 2021, from https://ccss.stjude.org/develop-a-study/gwas-data-resource/public-access-gwas-data-tables.html

Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., Sklar, P., Purcell (Leader), S. M., Stone, J. L., Sullivan, P. F., Ruderfer, D. M., McQuillin, A., Morris, D. W., O'Dushlaine, C. T., Corvin, A., Holmans, P. A., O'Donovan, M. C., Sklar, P., Wray, N. R., … Stanley Center for Psychiatric Research and Broad Institute of MIT and Harvard. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature, 460(7256), 748–752. https://doi.org/10.1038/nature08185

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., De Bakker, P. I., & Daly, M. J. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics, 81(3), 559–575.

Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-validation. Encyclopedia of Database Systems, 5, 532–538.

Robins, J. M., & Finkelstein, D. M. (2000). Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests. Biometrics, 56(3), 779–788. https://doi.org/10.1111/j.0006-341x.2000.00779.x

Robison, L. L. (2005). The Childhood Cancer Survivor Study: A resource for research of long-term outcomes among adult survivors of childhood cancer. Minnesota Medicine, 88(4), 45–49.

Robison, L. L., Armstrong, G. T., Boice, J. D., Chow, E. J., Davies, S. M., Donaldson, S. S., Green, D. M., Hammond, S., Meadows, A. T., Mertens, A. C., Mulvihill, J. J., Nathan, P. C., Neglia, J. P., Packer, R. J., Rajaraman, P., Sklar, C. A., Stovall, M., Strong, L. C., Yasui, Y., & Zeltzer, L. K. (2009). The Childhood Cancer Survivor Study: A National Cancer Institute–Supported Resource for Outcome and Intervention Research. Journal of Clinical Oncology, 27(14), 2308–2318. https://doi.org/10.1200/JCO.2009.22.3339

Rosenberg, N. A., Huang, L., Jewett, E. M., Szpiech, Z. A., Jankovic, I., & Boehnke, M. (2010). Genome-wide association studies in diverse populations. Nature Reviews Genetics, 11(5), 356–366.

Rossetti, R., Ferrari, I., Bonomi, M., & Persani, L. (2017). Genetics of primary ovarian insufficiency. Clinical Genetics, 91(2), 183–198. https://doi.org/10.1111/cge.12921

Rubin, D. B. (2004). Multiple imputation for nonresponse in surveys (Vol. 81). John Wiley & Sons.

Rudnicka, E., Kruszewska, J., Klicka, K., Kowalczyk, J., Grymowicz, M., Skórska, J., Pięta, W., & Smolarczyk, R. (2018). Premature ovarian insufficiency – aetiopathology, epidemiology, and diagnostic evaluation. Przeglad Menopauzalny, 17(3), 105–108. https://doi.org/10.5114/pm.2018.78550

Seibert, T., Fan, C., Wang, Y., Zuber, V., Karunamuni, R., Parsons, J., Eeles, R., Easton, D., Kote-Jarai, Z., Amin Al Olama, A., Garcia, S., Muir, K., Grönberg, H., Wiklund, F., Aly, M., Schleutker, J., Sipeky, C., Tammela, T., Nordestgaard, B., & Dale, A. (2018). Polygenic hazard score to guide screening for aggressive prostate cancer: Development and validation in large scale cohorts. BMJ, 360, j5757. https://doi.org/10.1136/bmj.j5757

Shah, D., & Nagarajan, N. (2014). Premature menopause – Meeting the needs. Post Reprod Health, 20(2), 62–68. https://doi.org/10.1177/2053369114531909

Sklar, C. A., Mertens, A. C., Mitby, P., Whitton, J., Stovall, M., Kasper, C., Mulder, J., Green, D., Nicholson, H. S., Yasui, Y., & Robison, L. L. (2006). Premature Menopause in Survivors of Childhood Cancer: A Report From the Childhood Cancer Survivor Study. JNCI: Journal of the National Cancer Institute, 98(13), 890–896. https://doi.org/10.1093/jnci/djj243

Spiegelhalter, D. J. (1986). Probabilistic prediction in patient management and clinical trials. Statistics in Medicine, 5(5), 421–433. https://doi.org/10.1002/sim.4780050506

Stats—Kids cancer care. (n.d.). Retrieved August 6, 2021, from https://www.kidscancercare.ab.ca/childhood-cancer/stats

Stevens, R. J., & Poppe, K. K. (2020). Validation of clinical prediction models: What does the "calibration slope" really measure? Journal of Clinical Epidemiology, 118, 93–99. https://doi.org/10.1016/j.jclinepi.2019.09.016

Steyerberg, E. W. (2019). Clinical prediction models. Springer.

Van Calster, B., McLernon, D. J., van Smeden, M., Wynants, L., Steyerberg, E. W., Bossuyt, P., Collins, G. S., Macaskill, P., McLernon, D. J., Moons, K. G. M., Steyerberg, E. W., Van Calster, B., van Smeden, M., Vickers, A. J., & On behalf of Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. (2019). Calibration: The Achilles heel of predictive analytics. BMC Medicine, 17(1), 230. https://doi.org/10.1186/s12916-019-1466-7

Van Calster, B., Nieboer, D., Vergouwe, Y., De Cock, B., Pencina, M. J., & Steyerberg, E. W. (2016). A calibration hierarchy for risk models was defined: From utopia to empirical data. Journal of Clinical Epidemiology, 74, 167–176.

van Dooren, M. F., Bertoli-Avellab, A. M., & Oldenburg, R. A. (2009). Premature ovarian failure and gene polymorphisms. Current Opinion in Obstetrics & Gynecology, 21(4), 313–317.

Wallace, W. H. B., Smith, A. G., Kelsey, T. W., Edgar, A. E., & Anderson, R. A. (2014). Fertility preservation for girls and young women with cancer: Population-based validation of criteria for ovarian tissue cryopreservation. The Lancet. Oncology, 15(10), 1129–1136. https://doi.org/10.1016/S1470-2045(14)70334-1

Willer, C. J., Li, Y., & Abecasis, G. R. (2010). METAL: Fast and efficient meta-analysis of genomewide association scans. Bioinformatics, 26(17), 2190–2191.

Wray, N. R., Goddard, M. E., & Visscher, P. M. (2007). Prediction of individual genetic risk to disease from genome-wide association studies. Genome Research, 17(10), 1520–1528.

Wray, N. R., Yang, J., Hayes, B. J., Price, A. L., Goddard, M. E., & Visscher, P. M. (2013). Pitfalls of predicting complex traits from SNPs. Nature Reviews Genetics, 14(7), 507–515. https://doi.org/10.1038/nrg3457

Younis, J. S. (2012). Ovarian aging and implications for fertility female health. Minerva Endocrinologica, 37(1), 41–57.

Yuan, Y., Su, W., & Zhu, M. (2015). Threshold-Free Measures for Assessing the Performance of Medical Screening Tests. Frontiers in Public Health, 3, 57. https://doi.org/10.3389/fpubh.2015.00057

Zhao, Z., Wang, H., Jessup, J. A., Lindsey, S. H., Chappell, M. C., & Groban, L. (2014). Role of estrogen in diastolic dysfunction. American Journal of Physiology-Heart and Circulatory Physiology, 306(5), H628–H640.

Appendices

Appendix A Data dictionary of variables

Table A-1 Data dictionary

	Variable names	Type (units)	Description
outcome	status	categorical/factor	ovarian status factor (4 levels): *Normal, AOF, PM, SPM*
	a_event	numeric (years)	age at event
demographics	age_dx	numeric (years)	age at diagnosis
	diagnose	categorical/factor	cancer diagnose type (8 levels): *Leukemia, CNS, HD, HNL, Kidney (Wilms), Neuroblastoma, Soft tissue sarcoma, Bone cancer*
	race	categorical/factor	race (3 levels): *White, Black, Other*
BMT	bmt	categorical/factor	received BMT within 5 years from first cancer diagnosis factor (2 levels): *No Yes*
radiotherapy	tbidose	numeric (cGy)	cumulative radiation doses to total body within 5 years
	minovary	numeric (cGy)	Minimum cumulative ovarian radiation dose within 5 years. The average radiation doses to right and left ovaries were estimated separately, and the lower dose was recorded as minimum ovarian radiation dose.
	pitdose	numeric (cGy)	cumulative radiation doses to pituitary within 5 years
Chemo therapy	bcnu	numeric (mg/m^2)	cumulative dose of BCNU within 5 years
	busulfan	numeric (mg/m^2)	cumulative dose of busulfan within 5 years
	ccnu	numeric (mg/m^2)	cumulative dose of CCNU within 5 years
	chlorambucil	numeric (mg/m^2)	cumulative dose of chlorambucil within 5 years
	cyclophosphamide	numeric (mg/m^2)	cumulative dose of cyclophosphamide within 5 years
	ifosfamide	numeric (mg/m^2)	cumulative dose of ifosfamide within 5 years
	melphalan	numeric (mg/m^2)	cumulative dose of melphalan within 5 years
	nitrogen_mustard	numeric (mg/m^2)	cumulative dose of nitrogen_mustard within 5 years
	procarbazine	numeric (mg/m^2)	cumulative dose of procarbazine within 5 years
	thiotepa	numeric (mg/m^2)	cumulative dose of thiotepa within 5 years
	carboplatin	numeric (mg/m^2)	cumulative dose of carboplatin within 5 years
	cis_platinum	numeric (mg/m^2)	cumulative dose of cis_platinum within 5 years
	bleomycin	numeric (mg/m^2)	cumulative dose of bleomycin within 5 years
	daunorubicin	numeric (mg/m^2)	cumulative dose of daunorubicin within 5 years
	doxorubicin	numeric (mg/m^2)	cumulative dose of doxorubicin within 5 years
	idarubicin	numeric (mg/m^2)	cumulative dose of idarubicin within 5 years
	methotrexate	numeric (mg/m^2)	cumulative dose of methotrexate within 5 years
	mitoxantrone	numeric (mg/m^2)	cumulative dose of mitoxantrone within 5 years
	vm_26	numeric (mg/m^2)	cumulative dose of VM 26 within 5 years

	vp_16	numeric (mg/m^2)	cumulative dose of VP 16 within 5 years

Appendix B The exclusion criteria established in a previous study for CCSS original cohort

Exclusion criteria: Long-term (≥5-year) female survivors who:

- Had missing menstrual history information in any questionnaires;
- Did not reach age 18 at their latest follow-up questionnaire;
- Whose menstrual status cannot be determined;
- Were exposed to a cranial or pituitary radiation dose higher than 30 Gy; (suspected pituitary dysfunction)
- Had a history of tumors in the hypothalamus or pituitary region;
- Had a history of Turner or Down's Syndrome;
- Had a secondary malignancy within five years of primary cancer diagnosis

Appendix C Explanatory data analysis

Cancer diagnoses

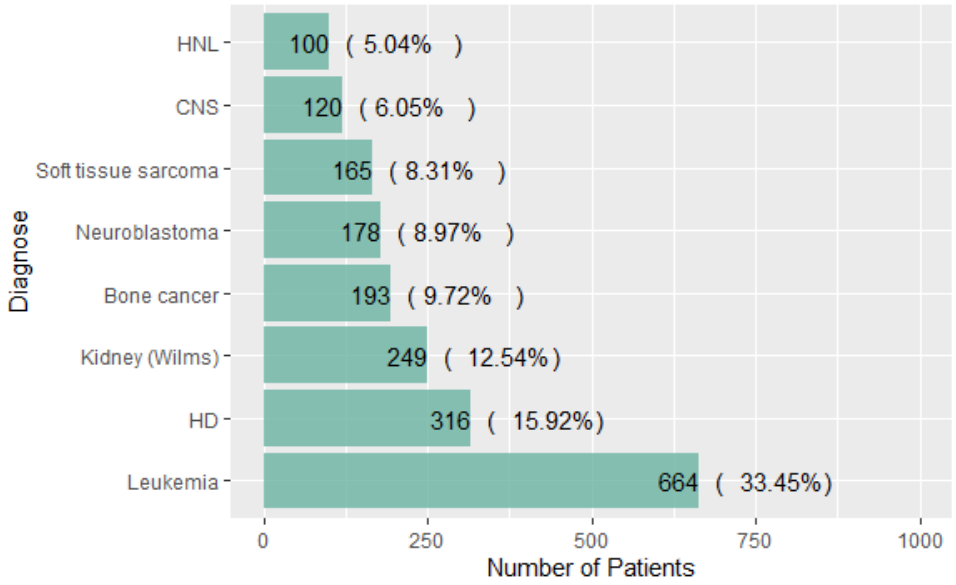


Figure C-1 The number of patients in different cancer diagnoses

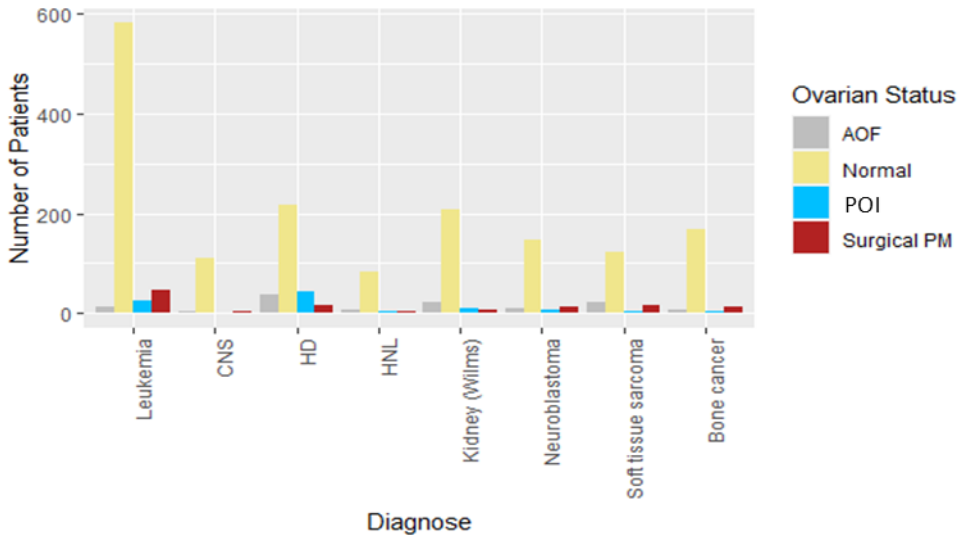


Figure C-2 Histogram of cancer diagnose by ovarian status

- Age at diagnose

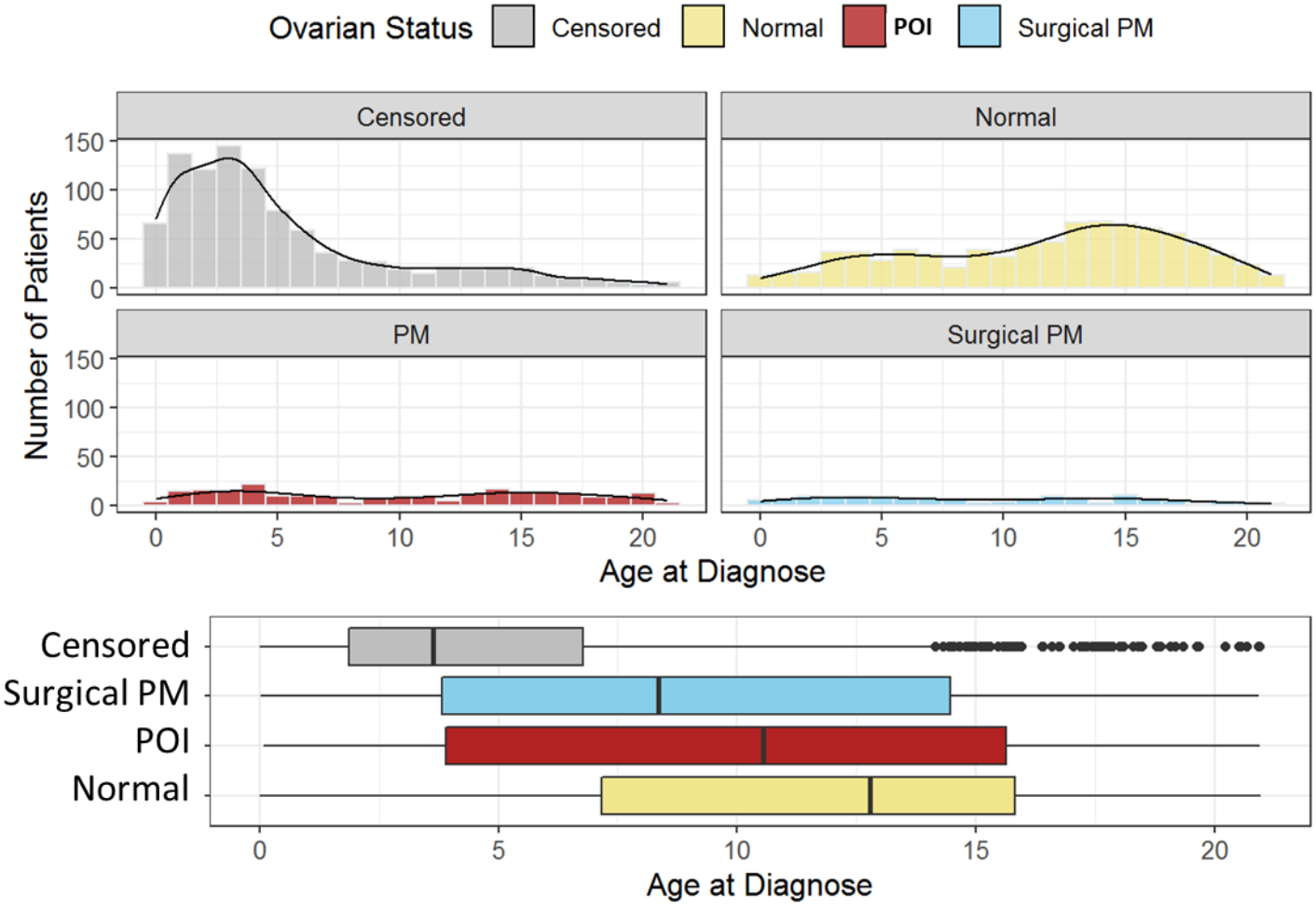


Figure C-3 Histogram and boxplot for age at cancer diagnosis by ovarian status

- Radiation

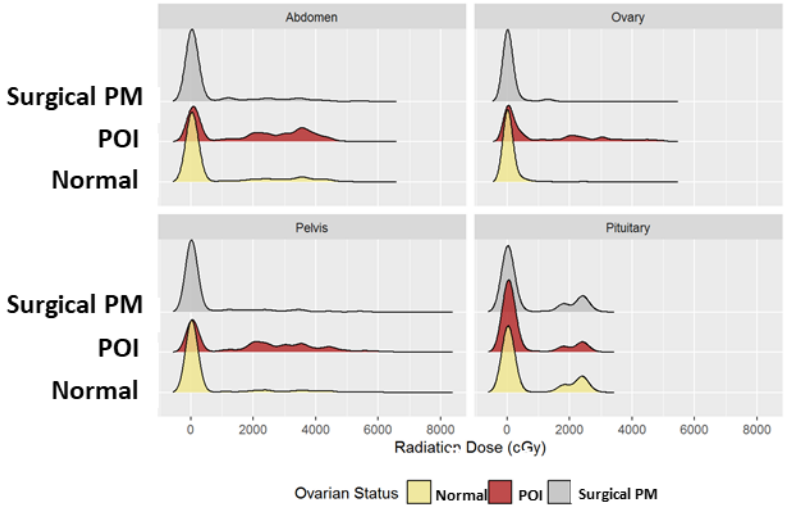


Figure C-4 Density plots of radiation dose for different body regions by ovarian status

Table C-1 The number and proportion of patients who received ovarian radiation therapy by ovarian status

	Censored (N=879)	Normal (N=762)	POI (N=226)	Surgical PM (N=118)	Overall (N=1985)
RT					
No	433 (49.3%)	261 (34.3%)	22 (9.7%)	41 (34.7%)	757 (38.1%)
Yes	446 (50.7%)	501 (65.7%)	204 (90.3%)	77 (65.3%)	1228 (61.9%

- Chemo-therapy agents

Table C-2 The number of patients who received chemo-agent by ovarian status

	No	Yes
Chemo Agent		
bcnu	1927	58
bleomycin	1891	94
busulfan	1984	1
carboplatin	1983	2
ccnu	1961	24
chlorambucil	1979	6
cis_platinum	1919	66
cyclophosphamide	1245	740
daunorubicin	1791	194
doxorubicin	1439	546
idarubicin	1985	0
ifosfamide	1975	10
melphalan	1961	24
methotrexate	1247	738
nitrogen_mustard	1880	105
procarbazine	1823	162
thiotepa	1981	4

- Age at event

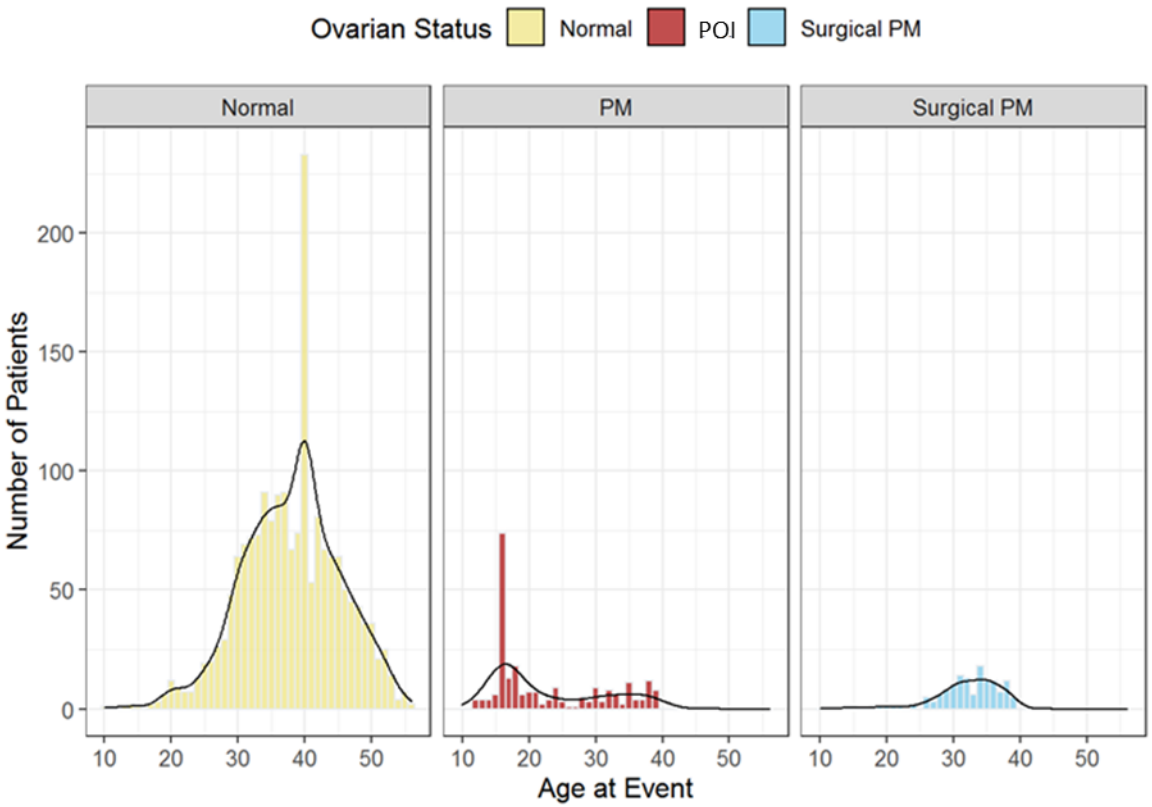


Figure C-5 The frequency plots for age at event by ovarian status

- Outcome status

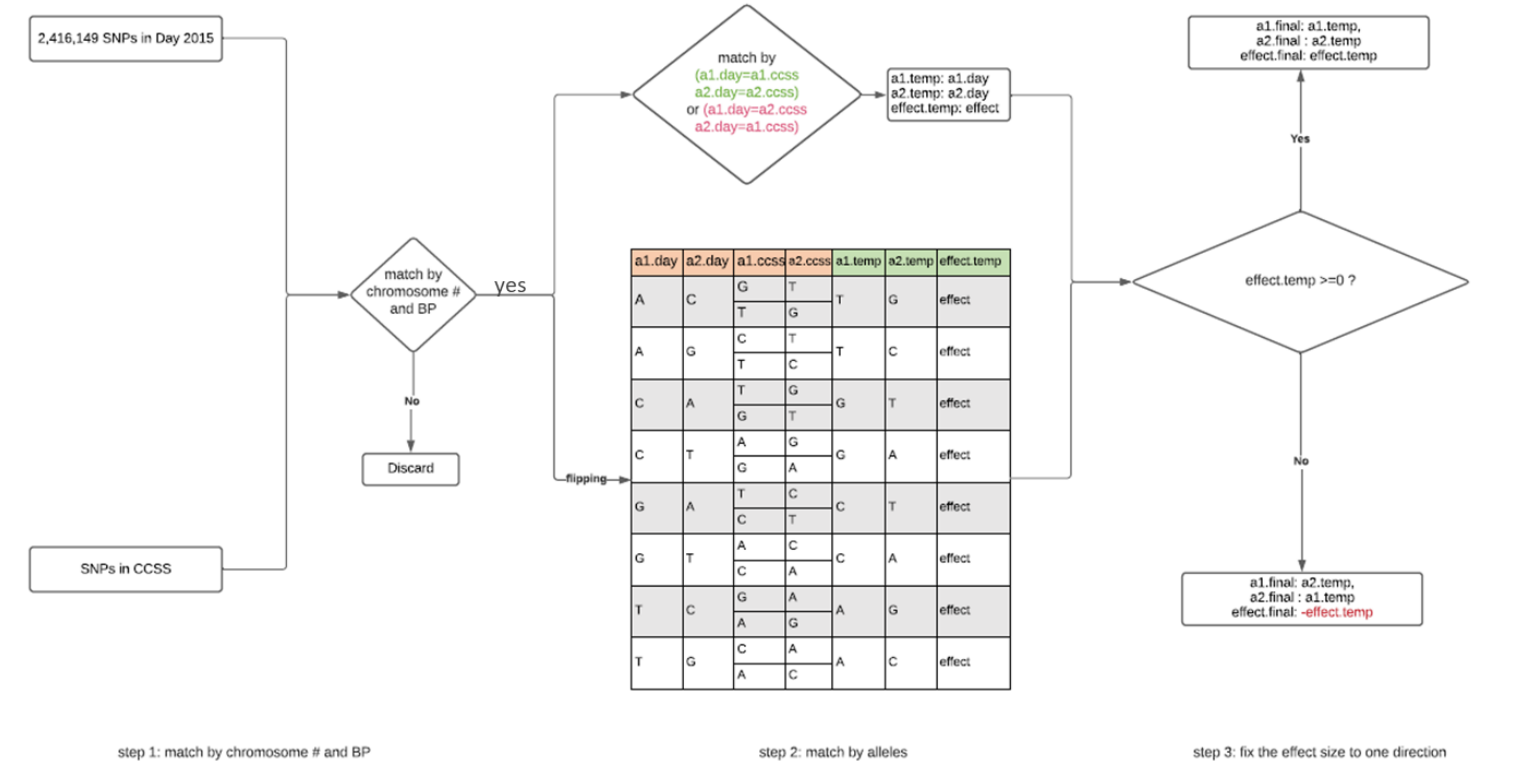
Table C-3 Summary statistics of ovarian status

Ovarian status	Number (proportion)
POI	226(11.4%)
Surgical PM	118 (5.9%)
Normal	1641 (82.7%)

Note: Surgical PM means surgical premature menopause

Appendix D Matching alleles between Day *et al.* and CCSS original cohort

The matching algorithm for the inclusion and exclusion of genetic variants used in Day *et al.*



step 1 original data	
Variable name	Description
a1.day	the effect allele in Day's study
a2.day	the reference allele in Day's study
effect	the effect size in Day's study
a1.ccss	the alleles in CCSS cohort data
a2.ccss	

step 2 temp data	
Variable name	Description
a1.temp	a temporary variable used to represent the effect allele
a2.temp	a temporary variable used to represent the reference allele
effect.temp	a temporary variable used to represent the effect size

step 3 final data	
Variable name	Description
a1.final	the effect allele that will be used as input in the PRS construction
a2.final	the final reference allele
effect.final	the effec size that will be used as input in the PRS construction

Figure D-1 Matching alleles between Day *et al.* and the CCSS original cohort

Appendix E Summary of the GWA studies for menopause-related phenotypes

Table E-1 Summary of extracted GWA studies in the general population/childhood cancer survivors

First author	Publication date	Study	Trait(s)	Discovery sample number and ancestry	Replication sample number and ancestry	Inclusion (yes/no) and reason
Watanabe K[1]	8/19/2019	A global overview of pleiotropy and genetic architecture in complex traits	age at menopause	119160 more than 80% are European	NA	**Yes**
Bae H[2]	6/10/2019	Genetic associations with age of menopause in familial longevity.	age at menopause	7611 European	3082 unknown ancestry	No, limted sample size
Horikoshi M[3]	5/17/2018	Elucidating the genetic architecture of reproductive ageing in the Japanese population.	age at menopause	43861 East Asian	32545 European	No, not of European ancestry in the discovery stage
Day FR[4]	11/1/2015	Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair.	age at menopause	69626 European	NA	**Yes**
Pyun JA[5]	9/16/2013	Genome-wide association studies and epistasis analyses of candidate genes related to age at menarche and age at natural menopause in a Korean population.	age at menopause	1827 East Asian	NA	No, not of European ancestry
Rahmani M[6]	4/16/2013	Shared genetic factors for age at natural menopause in Iranian and European women.	age at menopause	352 Greater Middle Eastern (Middle Eastern, North African or Persian)	573 Greater Middle Eastern (Middle Eastern, North African or Persian)38968 European	No, limited sample size; unknown ancestry
Ran S[7]	4/4/2013	Bivariate genome-wide association analyses identified genes with pleiotropic effects for femoral neck bone geometry and age at menarche.	age at menopause	1728 European	826 East Asian501 European	No, limited sample size

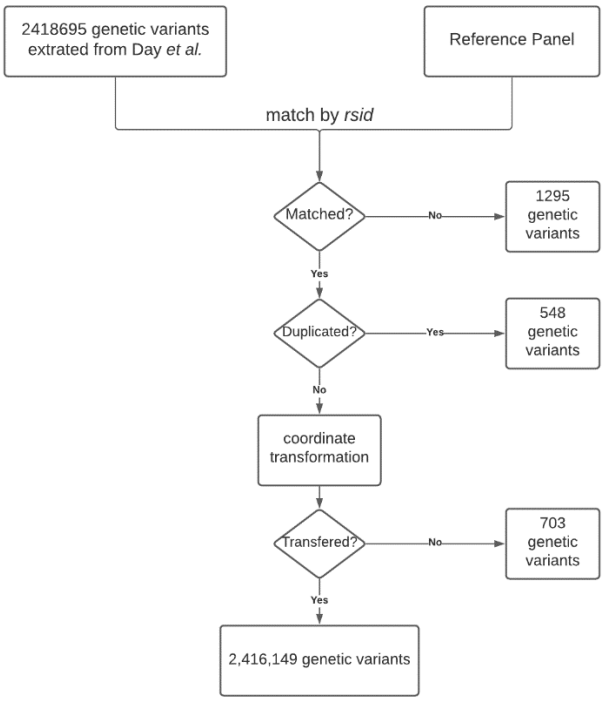
Perry JR[8]	1/9/2013	A genome-wide association study of early menopause and the combined impact of identified variants.	age at menopause	17091 European	8340 European	**Yes**
Stolk L[9]	1/22/2012	Meta-analyses identify 13 loci associated with age at menopause and highlight DNA repair and immune pathways.	age at menopause	38968 European	14435 European	**Yes**
Chen CT[10]	11/30/2011	Replication of loci influencing ages at menarche and menopause in Hispanic women: the Women's Health Initiative SHARe Study.	age at menarche, age at menopause	3468 Hispanic or Latin American	NA	No, not of European ancestry
He C[11]	5/17/2009	Genome-wide association studies identify loci associated with age at menarche and age at natural menopause.	age at menarche, age at menopause	17438 European	NA	**Yes**
Stolk L[12]	5/15/2009	Loci at chromosomes 13, 19 and 20 influence age at natural menopause.	age at menopause	2979 European	2560 European	No, limited sample size
Lunetta KL[13]	9/19/2007	Genetic correlates of longevity and selected age-related phenotypes: a genome-wide association study in the Framingham Study.	age at menopause, aging, exercise test, age at death	1345 unknown ancestry	NA	No, limited sample size; unknown ancestry
Park J[14]	10/26/2020	Association of an APBA3 Missense Variant with Risk of Premature Ovarian Failure in the Korean Female Population.	primary ovarian insufficiency	242 East Asian	322 East Asian	No, not of European ancestry
Brooke RJ[15]	2/8/2018	A High-risk Haplotype for Premature Menopause in Childhood Cancer Survivors Exposed to Gonadotoxic Therapy.	primary ovarian insufficiency	114 African unspecified 685 European	NA	No, limited sample size; mixed ancestry
Pyun JA[16]	2/10/2012	LAMC1 gene is associated with premature ovarian failure.	primary ovarian insufficiency	48 unknown ancestry	316 unknown ancestry	No, limited sample size; unkown ancestry
Qin Y[17]	10/18/2011	Association of 8q22.3 locus in Chinese Han with idiopathic premature ovarian failure (POF).	primary ovarian insufficiency	1286 East Asian	1200 East Asian	No, not of European ancestry
Knauff EA[18]	6/9/2009	Genome-wide association study in premature ovarian failure patients suggests ADAMTS19 as a possible candidate gene.	primary ovarian insufficiency	334 European	150 European	No, limited sample size

Note: sorted by trait and publication date; the superscript represents the ordered number.

Appendix F Quality control of genetic variants in Day *et al.* and CCSS original cohort

Coordinate Transformation

Among the three selected GWA studies, the genetic variants extracted from Day *et al*. and the CCSS original cohort are in different genome build, which requires special handling. The following flow chart illustrates the quality control of the genetic variants of the Day paper. Initially, around 2.5 million genetic variants were obtained from reproGen.[13] The missing reference allele information was imputed using a reference panel by matching the genetic identifier of genetic variants from GWA studies and the reference panel. 1295 unmatched and 548 duplicated genetic variants were excluded. The remaining 2,416,852 genetic variants were on hg 18 build, thus were further transformed into hg 19. The flowchart on the next page showed the workflow.



Note: Matched by rsid: rsid is the identifier of the genetic variants; Coordinate transformation: convert the coordinates of genetic variants in hg18 to hg19

Figure F-1 Flowchart for coordinate transformation

Matching Base and Target Data

Furthermore, the genetic variants from base and target data were matched, the overlapped genetic variants were kept. Finally, 2,416,147 genetic variants that exist in both the GWA studies and CCSS original cohort were kept for the development of PRS as the summary statistics information of the remaining two GWA studies was complete. The genetic variants were matched using the same algorithm mentioned above.

Appendix G Using the Metal tool to re-estimate the effect sizes of genetic variants in two GWA studies conducted in the general population

This study used two GWA studies in the general population to compute two *g*PRSs (PRS48 and PRS69). A meta-analysis was done to combine the evidence for association from these two GWA studies in the general population. The Metal tool can be used for this purpose. Briefly, the effect size estimates and standard error of the genetic variants from these two studies are used as input; each genetic variant will be assigned a weight, which is represented by:

$$w_i = 1/se^2$$

Where i refers to the ith genetic variant, and se refers to the standard error. The coefficient estimate and standard error for each variant is re-estimated with the weight:

$$se = \sqrt{1/\sum_i w_i}\ ,\ \beta = \sum_i \beta_i w_i \,/\, \sum_i w_i$$

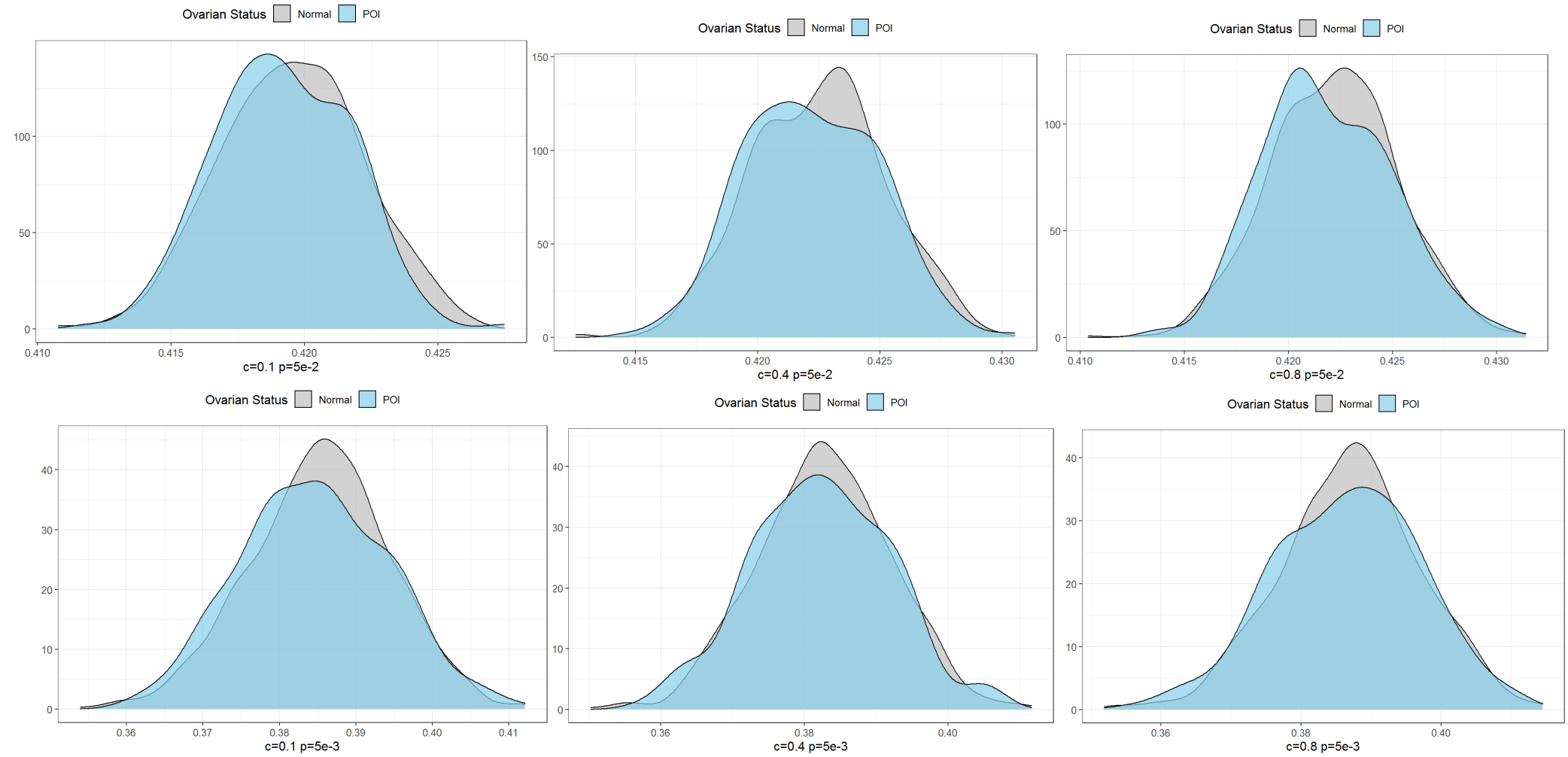
Furthermore, the Z-statistics and P-value can be derived from standard error and coefficient estimate, where $Z = \beta/se$ and $P = 2\Phi(|-Z|)$

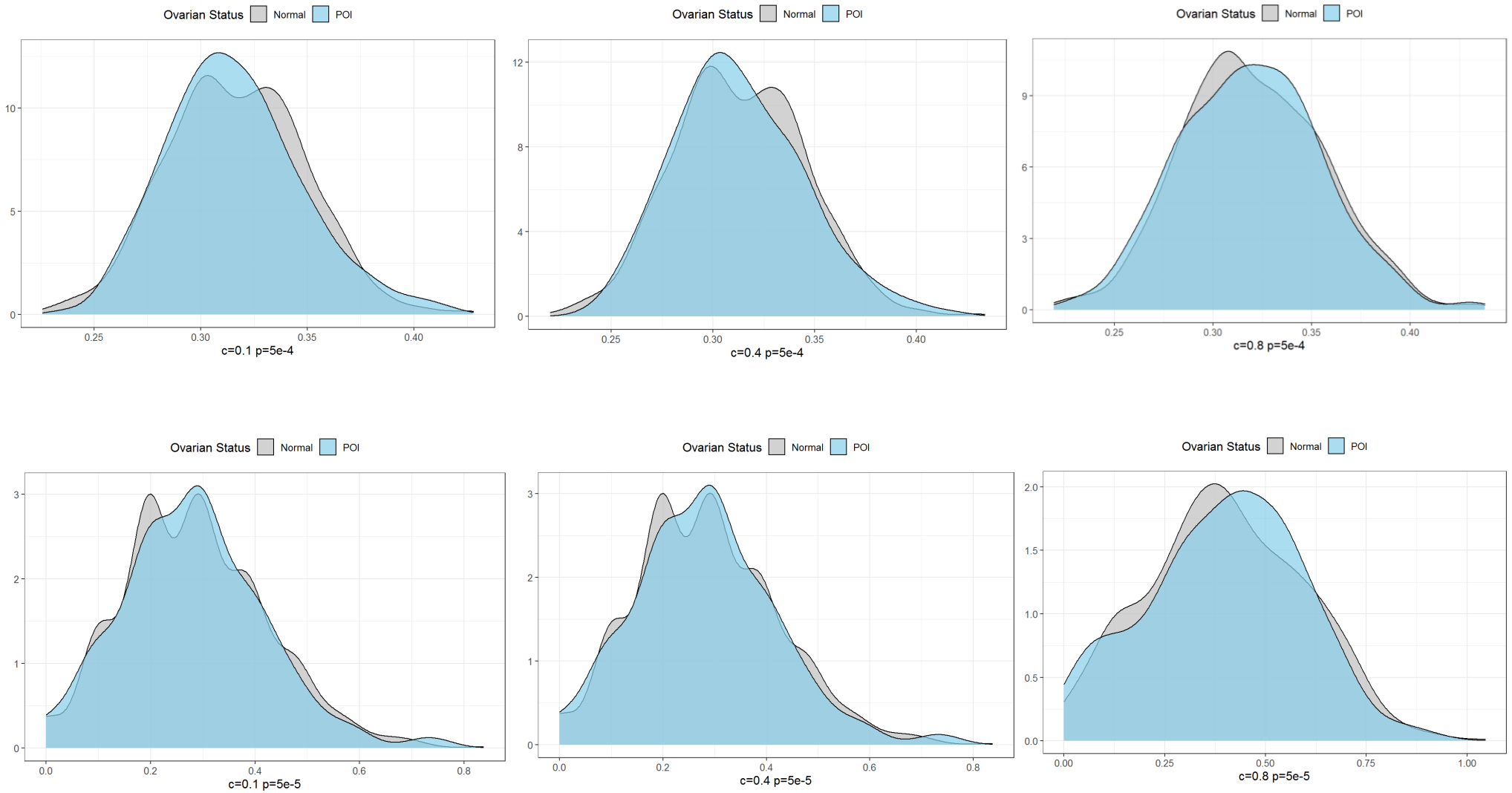
The summary statistics were re-estimated using Metal. A correlation coefficient of 0.1 and a P-value of 5e-08 were used as the threshold to select independent significant genetic variants for PRS construction. Finally, 262 genetic variants were selected to construct a new *g*PRS (named PRS262).

Appendix H Density plots of 12 candidate *ct*PRSs by ovarian status

The 12 candidate *ct*PRSs are listed, with the columns representing different P-value thresholds (5e-2, 5e-3, 5e-4, and 5e-5), and the rows representing the clumping parameters (0.1,0.4, and 0.8)

c=0.1 *c*=0.4 *c*=0.8





Note: The x-axis is the PRS value; the y-axis is the density.

Figure H-1 The density plots of 12 candidate *ct*PRSs

Appendix I Comparison of classification performance between the CRS and *g*PRS models

One of the Edinburgh Selection Criteria(14) for oocyte cryopreservation stated that patients with high risk (over 50%) of developing POI were suggested for ovarian tissue cryopreservation. Thus, 0.5 was used as a cutoff to compare the model performance further. Using .5 as the cutoff, some measurements were summarized in the following table: We see that most of the metrics were similar between CRS and *g*PRS models. However, with similar false-positive rates, the sensitivity of the *g*PRS model improved compared to the CRS model.

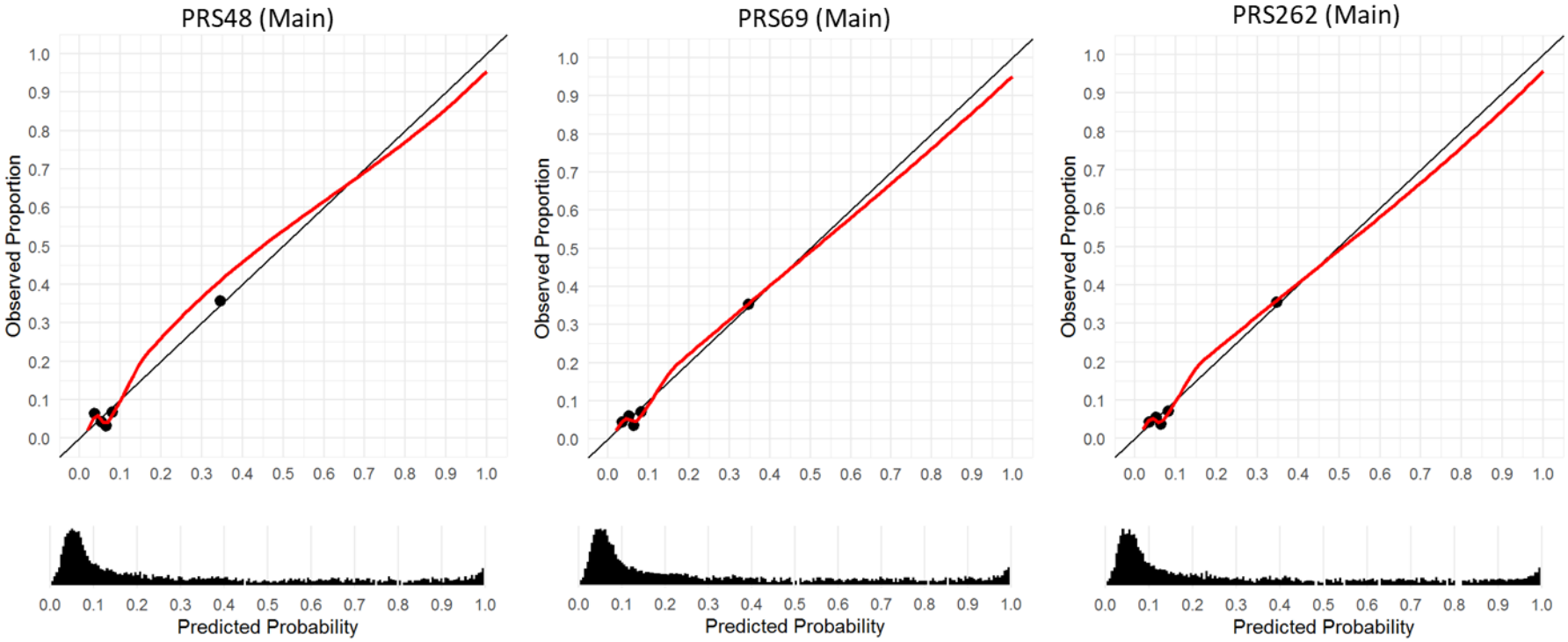
Table I-1 Some measurements of the models

			Sensitivity	Specificity	PPV	NPV
CRS			0.250	0.994	0.840	0.909
Main Effect Models		PRS69	**0.319**	0.991	0.822	**0.919**
		PRS6	**0.322**	0.990	0.809	**0.919**
		*ct*PRS	**0.324**	0.990	0.802	**0.920**
Interaction Models	CRS with	PRS69	**0.435**	0.971	0.656	**0.930**
		PRS6	**0.406**	0.974	0.667	**0.927**
		*ct*PRS	**0.395**	0.981	0.727	**0.927**
	RT with	PRS69	**0.338**	0.986	0.756	**0.921**
		PRS6	**0.335**	0.987	0.773	**0.921**
		*ct*PRS	**0.331**	0.989	0.791	**0.920**

Note: PPV is positive predictive value; NPV is the negative predictive value;

Appendix J Calibration and threshold-free performance metrics for three *g*PRSs

1) Calibration of *g*PRSs models (*g*PRSs: PRS48, PRS69, and PRS262)



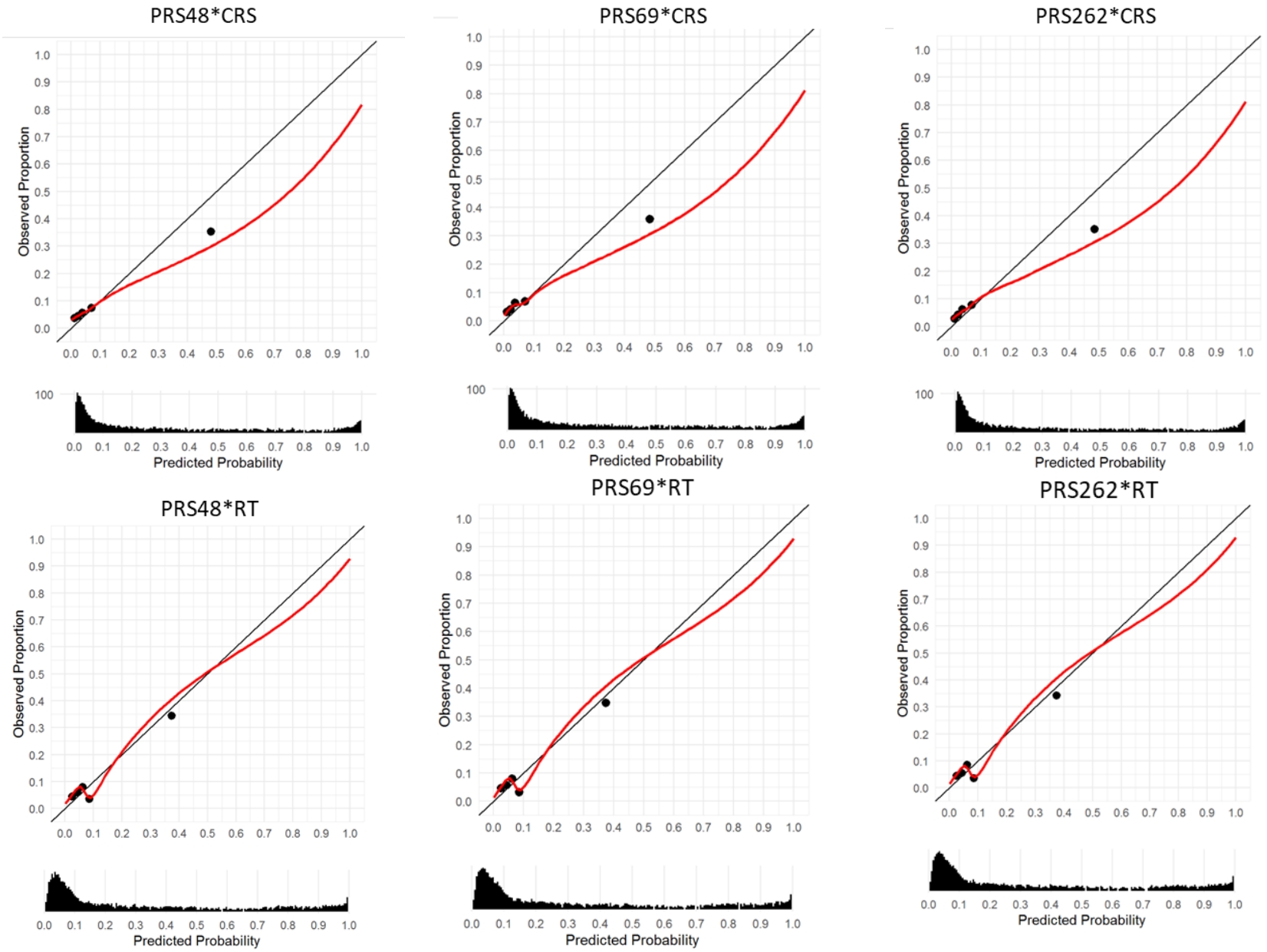


Figure J-1 Calibration plots of *g*PRS-based models

2) Threshold-free metrics for *g*PRS-based models

Table J-1 Threshold-free metrics for *g*PRS-based models

		sBrS	AP	AUC	Spiegelhalter-z
Main effect models	PRS48	0.274 (0.239, 0.307)	0.531(0.495, 0.568)	0.775 (0.748, 0.801)	0.185 (-1.775, 2.145)
	PRS69	0.277(0.242, 0.311)	0.532(0.495, 0.569)	0.780(0.756, 0.804)	**0.134(-1.826, 2.094)**
	PRS262	0.273 (0.239, 0.307)	0.529 (0.493, 0.566)	0.778 (0.752, 0.803)	0.209 (-1.751, 2.169)
Interaction models	PRS48*CRS	0.226 (0.179, 0.272)	0.532 (0.491, 0.573)	0.796 (0.776, 0.815)	6.557 (4.597, 8.517)
	PRS69*CRS	0.227 (0.18, 0.274)	0.537 (0.496, 0.578)	0.799 (0.78, 0.818)	6.686 (4.726, 8.646)
	PRS262*CRS	0.224 (0.175, 0.271)	0.534 (0.494, 0.576)	0.801 (0.781, 0.82)	6.889 (4.929, 8.849)
	PRS48*RT	0.266(0.228, 0.304)	0.519 (0.480, 0.556)	0.752 (0.731, 0.775)	1.473 (-0.487, 3.433)
	PRS69*RT	0.267 (0.229, 0.304)	0.520 (0.481, 0.557)	0.752 (0.731, 0.774)	1.485 (-0.475, 3.445)
	PRS262*RT	0.266 (0.229, 0.304)	0.519 (0.480, 0.556)	0.753 (0.731, 0.774)	1.509 (-0.451, 3.469)

Note: The table gives the pointe estimates and the corresponding 95% confidence intervals for each measurement

The results showed that:

- The gPRS based models generated similar performance regarding sBrS, AP and AUC.
- The Spiegelhalter-z statistics suggested the *g*PRS main effect models had better calibration, with PRS69 performed best.
- The PRS*RT models performed better than the PRS*CRS models

Appendix K Threshold-free performance metrics for 12 candidate *ct*PRSs

Table K-1 Model performance of twelve candidate *ct*PRSs

	*ct*PRS	sBrS	AP	AUC	Spiegelhalter-z	Averaged performance
Main effect models	c=0.1, t= 5e-02	0.275 (0.241, 0.309)	0.532 (0.495, 0.570)	0.784 (0.759, 0.81)	0.193 (-1.767, 2.153)	0.53
	c=0.1, t= 5e-03	0.274 (0.239, 0.308)	0.533 (0.496, 0.570)	0.795 (0.775, 0.815)	0.320 (-1.640, 2.280)	0.534
	c=0.1, t= 5e-04	0.277 (0.242, 0.310)	0.531 (0.494, 0.570)	0.781 (0.757, 0.805)	0.119 (-1.841, 2.079)	0.53
	c=0.1, t= 5e-05	0.276 (0.242, 0.310)	0.531 (0.495, 0.569)	0.776 (0.749, 0.801)	0.099 (-1.861, 2.059)	0.528
	c=0.4, t= 5e-02	0.273 (0.239, 0.308)	0.530 (0.494, 0.568)	0.778 (0.751, 0.804)	0.238 (-1.722, 2.198)	0.527
	c=0.4, t= 5e-03	0.274 (0.240, 0.308)	0.531 (0.493, 0.569)	0.790 (0.769, 0.810)	0.285 (-1.675, 2.245)	0.532
	c=0.4, t= 5e-04	0.276 (0.241, 0.310)	0.531 (0.494, 0.570)	0.779 (0.753, 0.803)	0.126 (-1.834, 2.086)	0.529
	c=0.4, t= 5e-05	0.276 (0.242, 0.310)	0.531 (0.495, 0.569)	0.776 (0.749, 0.801)	**0.099** (-1.861, 2.059)	0.528
	c=0.8, t= 5e-02	0.273 (0.239, 0.307)	0.531 (0.494, 0.568)	0.783 (0.758, 0.808)	0.288 (-1.672, 2.248)	0.529
	c=0.8, t= 5e-03	0.276 (0.242, 0.309)	0.535 (0.497, 0.571)	0.797 (0.777, 0.815)	0.243 (-1.717, 2.203)	0.536
	c=0.8, t= 5e-04	0.277 (0.242, 0.310)	0.533 (0.497, 0.571)	0.781 (0.756, 0.806)	0.149 (-1.811, 2.109)	0.53
	c=0.8, t= 5e-05	0.275 (0.24, 0.308)	0.531 (0.495, 0.568)	0.773 (0.746, 0.799)	0.13 (-1.830, 2.090)	0.526
Interaction models (with CRS)	c=0.1, t= 5e-02	0.224 (0.176, 0.271)	0.534 (0.493, 0.576)	0.8 (0.779, 0.819)	6.932 (4.972, 8.892)	0.519
	c=0.1, t= 5e-03	0.224 (0.176, 0.271)	0.535 (0.494, 0.576)	0.801 (0.781, 0.82)	6.968 (5.008, 8.928)	0.52
	c=0.1, t= 5e-04	0.228 (0.181, 0.275)	0.537 (0.497, 0.579)	0.801 (0.782, 0.822)	6.625 (4.665, 8.585)	0.522
	c=0.1, t= 5e-05	0.257 (0.214, 0.299)	0.528 (0.486, 0.572)	0.792 (0.773, 0.813)	**3.774 (1.814, 5.734)**	0.526
	c=0.4, t= 5e-02	0.224 (0.176, 0.271)	0.534 (0.493, 0.576)	0.799 (0.779, 0.819)	6.934 (4.974, 8.894)	0.519
	c=0.4, t= 5e-03	0.224 (0.176, 0.271)	0.535 (0.494, 0.576)	0.801 (0.78, 0.82)	6.952 (4.992, 8.912)	0.52
	c=0.4, t= 5e-04	0.229 (0.182, 0.275)	0.537 (0.497, 0.579)	0.8 (0.782, 0.821)	6.588 (4.628, 8.548)	0.522
	c=0.4, t= 5e-05	0.257 (0.214, 0.299)	0.528 (0.486, 0.572)	0.792 (0.773, 0.813)	**3.774 (1.814, 5.734)**	0.526
	c=0.8, t= 5e-02	0.224 (0.176, 0.271)	0.534 (0.493, 0.576)	0.8 (0.779, 0.819)	6.934 (4.974, 8.894)	0.519
	c=0.8, t= 5e-03	0.224 (0.176, 0.272)	0.535 (0.495, 0.577)	0.801 (0.781, 0.82)	6.948 (4.988, 8.908)	0.52
	c=0.8, t= 5e-04	0.231 (0.184, 0.277)	0.536 (0.495, 0.578)	0.8 (0.781, 0.82)	6.482 (4.522, 8.442)	0.522
	c=0.8, t= 5e-05	0.257 (0.215, 0.299)	0.528 (0.487, 0.572)	0.789 (0.769, 0.811)	**3.736 (1.776, 5.696)**	0.525

Note: Averaged performance is an average value of sBrS, AP and AUC.

The hyperparameter selection is based on the value of both the averaged performance and the Spiegelhalter-z statistics. A larger averaged performance and a lower Spiegelhalter-z statistic indicated better performance. The table showed that the averaged performances over different hyperparameter settings were similar. However, the Spiegelhalter-z statistic indicated that *ct*PRSs with a lower P-value performed better in the main effect and interaction models. A *ct*PRS with c=0.4 and t=5e-05 was selected and used for the PRS construction if given external data.

Appendix L Replication of two GWA studies

Using the CCSS samples as the replication cohort, time-specific logistic regression was done for each significant genetic variant reported in Watanabe et al. and Brooke et al. The following table summarized the direction of each variant-POI association and its corresponding P-value. Since the multiple imputation and cross-validation framework were used (see section 1.2.3), each genetic variant then had 25 coefficient estimates and corresponding P-values. The table summarized the total number of times that the association in the replication analysis is the same as that in the original GWA study over all training sets. The proportion that the association estimated from the replication cohort and the original GWAS is in the same direction. Besides, the total number of times that the associations are significant was also counted and reported. Similarly, the proportion of significant associations for each genetic variant was calculated.

The table showed that only a few genetic variants (14 out of 69) are statistically significant. Of the 69 genetic variants, 18 variants showed the same direction of association between the replication and the original GWA study among all training sets, with seven genetic variants showing significant associations sometimes.14 out of 69 genetic variants had the same directions between the replication and the original GWA studies 72-96 percent of the time. Eleven genetic variants always had a different direction of associations, with four variants showing significant associations sometimes.

Table L-1 The agreement between the replication analysis and Watanabe *et al.*

	genetic identifier	coefficient		P-value	
		# of negative coeffcients	proportion	# of significant coefficients	proportion
1	chr1.242011344.C.G_C	25	1	0	0
2	chr12.122203915.C.T_C	25	1	11	0.44
3	chr12.123593382.C.T_T	25	1	0	0
4	chr12.57146069.T.G_G	25	1	0	0
5	chr16.11948895.T.C_C	25	1	1	0.04
6	chr16.89786761.C.T_C	25	1	0	0
7	chr17.5327481.T.G_T	25	1	0	0
8	chr19.55827175.G.A_A	25	1	3	0.12
9	chr20.5580789.C.T_T	25	1	0	0
10	chr3.183573235.C.T_C	25	1	0	0
11	chr8.48926264.T.A_A	25	1	0	0
12	chr12.120189879.A.G_A	25	1	2	0.08
13	chr14.34985658.C.T_T	25	1	0	0
14	chr5.173410833.A.G_A	25	1	0	0
15	chr5.176425581.A.C_A	25	1	0	0
16	chr13.61113739.G.A_G	25	1	4	0.16
17	chr14.20938251.A.C_C	25	1	3	0.12
18	chr19.46890160.T.C_T	25	1	3	0.12
19	chr1.39361425.T.G_G	24	0.96	0	0
20	chr10.131590300.T.A_A	24	0.96	0	0
21	chr12.130822471.C.T_T	24	0.96	0	0
22	chr22.39021165.T.C_C	24	0.96	0	0
23	chr8.129607823.T.C_C	24	0.96	0	0
24	chr2.67597525.T.A_A	23	0.92	0	0
25	chr7.99785765.T.G_G	21	0.84	0	0
26	chr1.6701978.T.C_C	21	0.84	2	0.08
27	chr12.66704225.A.G_A	20	0.8	0	0
28	chr2.171794631.A.G_A	20	0.8	0	0
29	chr6.31601012.T.C_C	20	0.8	0	0
30	chr10.78008749.C.T_T	20	0.8	0	0
31	chr2.152280246.A.G_A	19	0.76	0	0
32	chr4.101069386.C.T_T	18	0.72	0	0
33	chr2.121146501.T.C_T	16	0.64	0	0
34	chr17.41245466.G.A_G	15	0.6	0	0
35	chr19.56321414.C.A_A	15	0.6	0	0
36	chr5.175953121.C.T_T	15	0.6	0	0
37	chr8.37884310.T.C_C	14	0.56	0	0
38	chr1.180961245.G.A_G	11	0.44	0	0
39	chr4.185745029.C.T_C	11	0.44	0	0
40	chr1.46728913.A.G_A	11	0.44	0	0
41	chr19.23166913.T.C_C	8	0.32	0	0

42	chr20.5948227.G.A_G	8	0.32	3	0.12
43	chr4.48814687.T.C_T	8	0.32	0	0
44	chr11.32549463.C.T_T	8	0.32	0	0
45	chr17.62479273.A.C_C	7	0.28	0	0
46	chr20.63244.A.C_C	7	0.28	0	0
47	chr15.41446950.C.T_T	6	0.24	0	0
48	chr15.86293503.G.C_G	6	0.24	0	0
49	chr17.55363674.A.T_T	5	0.2	0	0
50	chr20.25426173.C.T_C	5	0.2	0	0
51	chr7.56162172.G.C_C	5	0.2	0	0
52	chr10.126658075.G.A_A	4	0.16	0	0
53	chr10.97826334.A.G_G	3	0.12	0	0
54	chr11.30226356.T.C_T	3	0.12	0	0
55	chr12.10875928.A.C_A	1	0.04	0	0
56	chr12.12884357.A.G_A	1	0.04	0	0
57	chr12.76040392.T.C_C	1	0.04	5	0.2
58	chr15.89780538.A.G_G	1	0.04	0	0
59	chr16.35069526.G.A_A	0	0	0	0
60	chr17.37835240.C.T_C	0	0	1	0.04
61	chr2.27627366.G.A_A	0	0	15	0.6
62	chr2.48017768.C.A_C	0	0	0	0
63	chr4.84364808.T.C_T	0	0	0	0
64	chr5.154257868.T.C_T	0	0	0	0
65	chr5.6740468.T.G_G	0	0	20	0.8
66	chr6.10887276.C.G_G	0	0	0	0
67	chr7.105994726.G.A_A	0	0	0	0
68	chr7.5453537.G.T_G	0	0	0	0
69	chr9.33004375.C.A_A	0	0	2	0.08

Note: In the Watanabe *et al., the risk allele is positively related with the age at natural menopause. Therefore, the expected direction of the association should be negative when studying the association between genetic variant and POI.*

A similar table was created for the replication of the associations between POI and 6 genetic variants reported in the cancer survivor-based GWA study. However, the results differed from the original GWA study. Only one genetic variant (chr10.44103895.T.C_T) showed significant association, and the direction of the association changed. chr19.35619019.A.G_G and chr4.156116644.T.C_C showed same directions but the associations were not significant.

Table L-2 The agreement between the replication analysis and Brooke *et al.*

genetic identifier	coefficient		P-value	
	# of positive coefficients	proportion	# of significant coefficients	proportion
chr19.35619019.A.G_G	25	1	0	0
chr4.156116644.T.C_C	25	1	0	0
chr4.69830542.G.A_G	17	0.68	0	0
chr2.46000486.A.G_A	8	0.32	0	0
chr5.39416294.G.C_G	8	0.32	0	0
chr10.44103895.T.C_T	5	0.2	3	0.12

A summary of the coefficients and corresponding p-values for each variant are given in the following tables for your reference. Overall speaking, the direction of the coeffecients of 3 out of 6 genetic variants from the cancer survivor-based GWA study, and 32 out of 69 genetic variants for the general population-based GWA study were replicated respectively, though the associations were not significant.

Table L-3 Summary of the coefficient and corresponding P-values for genetic variants in Brooke *et al.*

	chr10.44103895.T.C_T	**chr19.35619019.A.G_G**	**chr2.46000486.A.G_A**	**chr4.156116644.T.C_C**	**chr4.69830542.G.A_G**	**chr5.39416294.G.C_G**
estimate						
Mean (SD)	-0.218 (0.148)	0.113 (0.076)	-0.0312 (0.060)	0.142 (0.070)	0.059(0.096)	-0.046 (0.072)
Median [Min, Max]	-0.282 [-0.389, 0.055]	0.087 [0.005, 0.300]	-0.016 [-0.142, 0.064]	0.121 [0.027, 0.303]	0.0721 [-0.092, 0.248]	-0.040 [-0.223, 0.062]
p-value						
Mean (SD)	0.291 (0.285)	0.583 (0.224)	0.780 (0.168)	0.468 (0.200)	0.696 (0.200)	0.750 (0.190)
Median [Min, Max]	0.147 [0.0320, 0.865]	0.626 [0.189, 0.974]	0.838 [0.448, 0.993]	0.529 [0.102, 0.872]	0.734 [0.260, 0.988]	0.800 [0.265, 0.998]

Table L-4 Summary of the coefficient and corresponding P-values for genetic variants in Watanabe *et al.*

genetic identifier	coefficient		P-value	
	Mean (SD)	Median [Min, Max]	Mean (SD)	Median [Min, Max]
chr1.180961245.G.A_G	0.017(0.097)	0.054 [-0.195, 0.126]	0.660 (0.174)	0.671 [0.307, 0.917]
chr1.242011344.C.G_C	-0.123 (0.077)	-0.107 [-0.316, -0.013]	0.544 (0.232)	0.558 [0.105, 0.941]
chr1.39361425.T.G_G	-0.196 (0.146)	-0.254 [-0.429, 0.114]	0.307 (0.248)	0.178 [0.066, 0.964]
chr1.46728913.A.G_A	0.044 (0.059)	0.062 [-0.0766, 0.135]	0.733 (0.105)	0.730 [0.521, 0.925]

chr1.6701978.T.C_C	-0.0348(0.088)	-0.046 [-0.177, 0.139]	0.659 (0.192)	0.659 [0.300, 0.991]
chr2.121146501.T.C_T	-0.008(0.115)	0.021 [-0.199, 0.175]	0.639 (0.234)	0.673 [0.277, 0.944]
chr2.152280246.A.G_A	-0.057 (0.104)	-0.064 [-0.235, 0.076]	0.646 (0.208)	0.693 [0.256, 0.995]
chr2.171794631.A.G_A	-0.068(0.119)	-0.065 [-0.246, 0.155]	0.575 (0.235)	0.600 [0.219, 0.980]
chr2.27627366.G.A_A	0.137 (0.088)	0.109 [0.031, 0.306]	0.506 (0.243)	0.564 [0.107, 0.874]
chr2.48017768.C.A_C	0.196 (0.046)	0.196 [0.131, 0.297]	0.385 (0.107)	0.371 [0.188, 0.616]
chr2.67597525.T.A_A	-0.052(0.091)	-0.0630 [-0.235, 0.138]	0.784 (0.132)	0.768 [0.485, 0.998]
chr3.183573235.C.T_C	-0.127 (0.052)	-0.135 [-0.239, -0.018]	0.553 (0.172)	0.494 [0.229, 0.931]
chr4.101069386.C.T_T	0.005 (0.070)	-0.009 [-0.114, 0.128]	0.802 (0.149)	0.831 [0.547, 0.995]
chr4.185745029.C.T_C	0.027 (0.055)	0.0218 [-0.085, 0.112]	0.797 (0.151)	0.783 [0.539, 0.992]
chr4.48814687.T.C_T	0.126 (0.132)	0.181 [-0.179, 0.258]	0.355 (0.151)	0.319 [0.154, 0.607]
chr4.84364808.T.C_T	0.237 (0.095)	0.273 [0.036, 0.360]	0.246 (0.236)	0.110 [0.043, 0.836]
chr5.154257868.T.C_T	0.216 (0.109)	0.191 [0.045, 0.436]	0.545 (0.198)	0.568 [0.186, 0.889]
chr5.173410833.A.G_A	-0.267 (0.170)	-0.293 [-0.566, 0.018]	0.343 (0.314)	0.181 [0.006, 0.935]
chr5.175953121.C.T_T	0.020 (0.058)	0.038 [-0.083, 0.122]	0.770 (0.106)	0.791 [0.489, 0.988]
chr5.176425581.A.C_A	-0.163 (0.105)	-0.159 [-0.315, 0.043]	0.410 (0.285)	0.395 [0.071, 0.920]
chr5.6740468.T.G_G	0.096(0.053)	0.098 [0.022, 0.198]	0.620 (0.195)	0.680 [0.290, 0.897]
chr6.10887276.C.G_G	0.327 (0.086)	0.313 [0.203, 0.525]	0.237 (0.105)	0.239 [0.060, 0.447]
chr6.31601012.T.C_C	-0.055 (0.089)	-0.065 [-0.204, 0.084]	0.703 (0.201)	0.736 [0.357, 0.993]
chr7.105994726.G.A_A	0.231 (0.103)	0.229 [0.067, 0.447]	0.278 (0.191)	0.264 [0.018, 0.721]
chr7.5453537.G.T_G	0.299 (0.058)	0.306 [0.179, 0.412]	0.148 (0.0798)	0.129 [0.035, 0.381]
chr7.56162172.G.C_C	0.105 (0.065)	0.103 [-0.009, 0.242]	0.607 (0.220)	0.591 [0.212, 0.979]
chr7.99785765.T.G_G	-0.060 (0.091)	-0.093 [-0.199, 0.125]	0.674 (0.180)	0.650 [0.362, 0.975]
chr8.129607823.T.C_C	-0.055 (0.0601)	-0.050 [-0.156, 0.064]	0.743 (0.148)	0.801 [0.457, 0.976]
chr8.37884310.T.C_C	0.060 (0.126)	0.073 [-0.163, 0.259]	0.608 (0.223)	0.560 [0.261, 0.992]
chr8.48926264.T.A_A	-0.482 (0.186)	-0.522 [-0.786, -0.059]	0.253 (0.243)	0.172 [0.019, 0.897]
chr9.33004375.C.A_A	0.223 (0.087)	0.207 [0.093, 0.366]	0.280 (0.192)	0.275 [0.037, 0.696]
chr10.126658075.G.A_A	0.170 (0.039)	0.155 [0.113, 0.263]	0.360 (0.102)	0.386 [0.145, 0.513]
chr10.131590300.T.A_A	-0.081 (0.085)	-0.071 [-0.266, 0.027]	0.665 (0.255)	0.684 [0.157, 0.972]
chr10.78008749.C.T_T	-0.035 (0.117)	-0.021 [-0.284, 0.128]	0.654 (0.226)	0.693 [0.150, 0.932]
chr10.97826334.A.G_G	0.789 (0.167)	0.725 [0.573, 1.18]	0.0723 (0.055)	0.068 [0.001, 0.175]
chr11.30226356.T.C_T	0.299 (0.105)	0.283 [0.145, 0.593]	0.330 (0.137)	0.323 [0.062, 0.589]
chr11.32549463.C.T_T	0.119 (0.096)	0.153 [-0.0273, 0.271]	0.530 (0.309)	0.368 [0.120, 0.999]
chr12.10875928.A.C_A	0.326 (0.158)	0.373 [0.064, 0.565]	0.465 (0.237)	0.364 [0.175, 0.890]
chr12.120189879.A.G_A	-0.119 (0.054)	-0.108 [-0.232, 0.006]	0.524 (0.178)	0.563 [0.196, 0.971]
chr12.122203915.C.T_C	-0.198 (0.128)	-0.131 [-0.409, -0.044]	0.431 (0.282)	0.544 [0.045, 0.826]
chr12.123593382.C.T_T	-0.449 (0.178)	-0.505 [-0.656, -0.109]	0.143 (0.222)	0.039 [0.008, 0.630]
chr12.12884357.A.G_A	0.255 (0.084)	0.273 [0.081, 0.390]	0.226 (0.178)	0.176 [0.034, 0.673]
chr12.130822471.C.T_T	-0.039 (0.113)	-0.078 [-0.192, 0.183]	0.577 (0.208)	0.562 [0.261, 0.947]
chr12.57146069.T.G_G	-0.228 (0.088)	-0.246 [-0.405, -0.049]	0.462 (0.179)	0.438 [0.168, 0.868]
chr12.66704225.A.G_A	-0.255 (0.338)	-0.151 [-0.897, 0.113]	0.566 (0.371)	0.708 [0.013, 0.973]
chr12.76040392.T.C_C	0.270 (0.162)	0.261 [0.022, 0.626]	0.493 (0.242)	0.464 [0.081, 0.949]
chr13.61113739.G.A_G	-0.124 (0.101)	-0.152 [-0.263, 0.048]	0.514 (0.298)	0.412 [0.150, 0.995]
chr14.20938251.A.C_C	-0.086 (0.062)	-0.089 [-0.189, 0.0416]	0.675 (0.182)	0.689 [0.382, 0.995]
chr14.34985658.C.T_T	-0.691 (0.475)	-0.606 [-1.99, 0.058]	0.412 (0.222)	0.366 [0.062, 0.913]
chr15.41446950.C.T_T	0.065 (0.034)	0.073 [-0.004, 0.131]	0.717 (0.143)	0.679 [0.465, 0.997]
chr15.86293503.G.C_G	0.077 (0.046)	0.071 [-0.004, 0.165]	0.678 (0.177)	0.694 [0.339, 0.977]

chr15.89780538.A.G_G	0.238 (0.068)	0.219 [0.131, 0.374]	0.209 (0.113)	0.195 [0.059, 0.444]
chr16.11948895.T.C_C	-0.122 (0.067)	-0.093 [-0.274, -0.055]	0.532 (0.197)	0.614 [0.113, 0.767]
chr16.35069526.G.A_A	0.237 (0.106)	0.278 [0.007, 0.348]	0.251 (0.300)	0.102 [0.042, 0.966]
chr16.89786761.C.T_C	-0.187 (0.093)	-0.173 [-0.366, -0.038]	0.397 (0.228)	0.373 [0.075, 0.848]
chr17.37835240.C.T_C	0.159 (0.088)	0.185 [0.008, 0.302]	0.449 (0.252)	0.338 [0.122, 0.964]
chr17.41245466.G.A_G	0.023 (0.074)	0.032 [-0.091, 0.137]	0.739 (0.131)	0.721 [0.505, 0.982]
chr17.5327481.T.G_T	-0.483 (0.124)	-0.526 [-0.655, -0.232]	0.0421 (0.075)	0.006 [0.001, 0.241]
chr17.55363674.A.T_T	0.194 (0.0824)	0.212 [-0.007, 0.316]	0.339 (0.228)	0.257 [0.084, 0.975]
chr17.62479273.A.C_C	0.188 (0.134)	0.217 [-0.086, 0.371]	0.317 (0.270)	0.239 [0.034, 0.971]
chr19.23166913.T.C_C	0.169 (0.176)	0.235 [-0.168, 0.371]	0.350 (0.186)	0.315 [0.112, 0.835]
chr19.46890160.T.C_T	-0.080 (0.077)	-0.062 [-0.237, 0.030]	0.726 (0.221)	0.791 [0.312, 0.998]
chr19.55827175.G.A_A	-0.094 (0.067)	-0.109 [-0.242, -0.003]	0.641 (0.238)	0.597 [0.212, 0.984]
chr19.56321414.C.A_A	0.0386 (0.143)	0.024 [-0.208, 0.320]	0.755 (0.177)	0.792 [0.386, 1.00]
chr20.25426173.C.T_C	0.157 (0.117)	0.135 [-0.011, 0.402]	0.705 (0.204)	0.737 [0.310, 0.989]
chr20.5580789.C.T_T	-0.329 (0.172)	-0.277 [-0.668, -0.073]	0.615 (0.159)	0.635 [0.332, 0.906]
chr20.5948227.G.A_G	0.171 (0.183)	0.156 [-0.140, 0.449]	0.624 (0.188)	0.683 [0.271, 0.832]
chr20.63244.A.C_C	0.129 (0.081)	0.149 [-0.010, 0.279]	0.583 (0.242)	0.509 [0.188, 0.993]
chr22.39021165.T.C_C	-0.007 (0.077)	-0.028 [-0.111, 0.195]	0.784 (0.170)	0.803 [0.428, 0.998]