

Risk Prediction of Primary Ovarian Insufficiency in Childhood Cancer Survivors

Lin Yu

2021-08-07

Contents

Introduction	5
1 Literature Review	7
1.1 Background	7
1.2 Methodology Review	10
2 Polygenic Risk Score Construction	13
3 Risk Prediction for Primary Ovarian Insufficiency in Childhood Cancer Survivors	15
4 Conclusion	17
5 Final Words	19

Introduction

Childhood cancer survivors are a rapidly growing group in developed countries due to the advancement in cancer treatments¹. 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 82% in Canada. This is a significant improvement, and consequently, the size of the childhood cancer survivor population has grown dramatically, to 300,000 individuals in Canada. However, the remarkable increases in survival have been accompanied with adverse effects later in life known as late effects. 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.

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. POI can occur early, during, or immediately following the completion of cancer treatment or, more commonly, in the years that follow the completion of cancer treatments but prior to age 40. In the general population, the prevalence of POI is approximately 1% whereas a study reported that 10.9% (100 out of 921) of childhood cancer survivors developed POI. Additionally, another study reported a 13-fold (95% CI = 3.26 to 53.51; $P < .001$) increased risk in developing POI after five years of cancer diagnosis compared to their healthy siblings.

The negative impacts of POI, including infertility, an increased risk of chronic disease, and a reduction in overall life quality, have prompted researchers 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. In recent decades, Genome-wide association studies (GWAS) have identified low-risk variants related to menopause-related phenotypes. However, little is known regarding variations' potential to identify POI at different risk levels, and the effect of gene-treatment interactions on POI. Therefore, this study aims to develop prediction models using genetic information from GWAS studies in combination with clinical risk factors, and investigate potential gene-treatment

interactions.

The purpose of the predictive models is to identify individuals who are at high risk of POI. Thus timely and appropriate interventions can be taken. For individuals at high risk of developing POI, patients can be counseled before or shortly after the cancer treatment regarding the risk and the need for fertility preservation such as oocyte and ovarian tissue cryopreservation[ref]. If the risk of developing POI is low, clinicians can provide consoles to patients and their families and 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 survivorship of female cancer survivors and to ensure that appropriate interventions are taken to improve the well being of female cancer survivors.

This thesis is structured as follows. The remainder of Chapter 1 reviews the literature on primary ovarian insufficiency, previous related work and genetics for POI, describes the statistical methods used in model development and evaluation, followed by an introduction to the data. Chapter 2 presents the construction of polygenic risk scores. Chapter 3 highlights the model development and evaluation. Finally, Chapter 4 summarizes the findings, discusses study limitations, and provides recommendations for future research.

Chapter 1

Literature Review

1.1 Background

It is estimated that approximately 90% of women experience menopause between the age of 45 and 55 years [ref], with the median age at natural menopause occurring between 50 and 52 years of age [ref]. 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 [ref]. 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, which accounts for about 10% of cancer survivors [ref].

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 postmenopause years [ref]. Lower levels of estrogen following menopause also increase the risk of developing hypertension and ventricular remodeling [ref]. At the same time, menopause along with the 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 [ref].

Fertility Preservation

Another 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 [ref]. However, a study has suggested that cancer patients feel difficult to make decisions about fertility preservation [ref]. 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 [ref]. Additionally, uncertainty makes

it challenging to make decisions[ref]. 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 [ref]. Some decision aid methods have been proposed [ref] , which mainly focused on improving patients' knowledge about fertility preservation.

Clinical Risk Factors and Risk Prediction Model

Extensive studies have been undertaken to identify treatment-related risk factors associated with compromised reproductive function following cancer treatment[ref]. Chemotherapy agents, especially alkylating agents (such as busulfan, cyclophosphamide, lomustine, and 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 [ref] . Also, radiation to the ovary, abdominal or pelvic sites 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 [ref] .

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 [ref], radiation exposure to the ovary, abdomen and pelvis, age at cancer diagnosis, and hematopoietic stem-cell transplant receipt to build models to predict individual's risk for developing menopause within five years following the time of cancer diagnosis. Zhe has investigated the association between clinical risk factors, their potential interactions on POI, and incorporated them into a predictive model for age-specific POI risk in female cancer survivors. [ref]

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.[ref] 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.[ref] For example, Rajkovic et al. discovered that the Newborn ovary homeobox gene (NOBOX) plays a critical role in early folliculogenesis [ref]. 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 [ref]. 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 [ref]. So far, more than sixty genetic variants have been identified, accounting for 10–13% of the variance of POI in the general population [ref]. A list of several genes that affect the development and maturation of POI is included in the appendix

Genetic Variants and Genome-wide Association Study (GWAS)

DNA sequence is formed from a chain of four nucleotide bases: A, C, G and T. All humans have near-identical DNA sequences across the estimated six billion-letter codes for their genome¹³. However, slight differences exist between individuals. If more than 1% of a population does not carry the same nucleotide at a specific position within the DNA sequence, then this variation can be classified as a single nucleotide polymorphism (SNP) or genetic variant¹⁴.

Many techniques were applied to discover the genetic variants for different phenotypes. A popular technique, Genome-wide Association Study (GWAS) has been widely conducted to study the genetics of phenotypes in recent decades¹¹, which aims to find common variants¹². GWAS technique consists of screening and comparing genetic variants in patients with the disease of interest (or phenotype) and healthy controls. In GWAS, the association between a genetic variant to the phenotype is observed when there is a statistically significant difference of the genetic variant between diseased patients and healthy controls.

GWAS studies have been conducted in several ancestries. As of 2019, 78% of the GWAS studies were performed in the European ancestry, 10% were conducted in Asian ancestry. For POI trait, there exist five GWAS studies to date, all five studies have limited sample size, with the largest sample size being around 1300. four of them were performed in the general population in either European or East Asian ancestries, and the remaining one was conducted in the childhood cancer survivors. a summary of the studies was provided in the appendix the table from cat log more GWASs studies with a larger study sample size were available for age at natural menopause. 14 GWAS studies are found and the largest one is Day et al, with #69626 individuals of European ancestry.

GWAS data for Age at Natural Menopause in the General Population Ideally, the GWAS information of POI is the perfect data source for studying the association between common variants and POI. However, the existing GWAS data for POI in the general population was limited by the small total study population sample size of the GWAS studies. Instead, this study focused on the GWAS data of age at natural menopause in the general population, as it has been hypothesized that age at natural menopause and POI are possibly manifestations of the same underlying genetic susceptibility owing to the inheritance patterns observed. [ref] Studies also showed that POI and age at natural menopause share common genetic factors involved in DNA repair and maintenance [ref]. The following describes some main findings of the two most recent GWAS studies with a reasonable sample size for age at natural menopause in the general population.

A meta-analysis of GWAS conducted in the general population analyzed up to 69,360 women of European ancestry. 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 associated with age at menopause, 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.[ref]

European population PRS69 [TO BE FINISHED]

GWAS data for Primary Ovarian Insufficiency in the Cancer Survivors

The genetics of menopause in the cancer survivor population is less investigated. To date, only Brooke et al. have conducted a GWAS study in the cancer survivor population.[ref] The study conducted GWAS to identify genetic variants associated with clinically diagnosed POI in 799 female cancer survivors in the St. Jude Lifetime Cohort Study (SJLIFE). The analyses were adjusted for cyclophosphamide equivalent dose of alkylating agents and ovarian radiotherapy (RT) dose. 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 seen in more than 50% of POI patients and 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 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).

Potential of Polygenic Risk Score

Although the increasing use of GWAS has led to the identification of many variants associated with POI, the effect size of genetic variants identified from GWAS are typically small and account for only a small fraction of association [ref], meaning that single variants have limited predictive power[ref]. Thus, this study evaluated the genetic risk for POI in the form of a polygenic risk score (PRS), or a score that combines the estimated effects of many disease-associated genetic variants reported in published GWAS data. PRSs have been proposed as a genetic risk prediction tool for a wide range of diseases [ref]; According to PGS catalog, an open database of published polygenic risk scores, 829 polygenic risk scores were built over 214 traits as of July 2021. A clinically-useful PRS would allow clinicians to identify individuals at elevated risk of disease, thus informing disease screening, therapeutic interventions, and life planning to prevent or delay the onset of disease.

1.2 Methodology Review

1.2.1 Polygenic Risk Score

A polygenic risk score or PRS (also termed a polygenic score or genetic risk score) estimates an individual's genetic liability to a trait or disease.[ref] PRS is derived from genotype profile and relevant genome-wide association study

(GWAS) data. PRS is a sum of genome-wide genotypes, weighted by corresponding genotype effect size estimates derived from GWAS summary statistics. The general steps for computing PRS include selecting base and target data, quality control of base and target data, finally, the construction of PRS.

Base and Target Data

Two key input data sets are required to characterize PRS based on the definition. The base data (GWAS studies), consisting of summary statistics of the association between genetic variants and phenotypes, are needed to characterize PRS. Target data, usually new samples independent of the base samples and with genotype and phenotype data for each individual, are used to perform the PRS analysis.

The base data (i.e., summary statistics of GWAS) are usually publicly available. the following summary statistic information is needed for the PRS construction:

Selection of GWAS

Several aspects were considered when it came to the selection of GWAS studies, including the sample size of the GWAS, the ancestry of the samples, the quality control process, and the phenotypes being studied in GWAS.

Sample Size of GWAS: Given that the PRS is built based on the summary statistics from GWAS studies, the GWAS 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. GWAS studies with a sample size larger than 10,000 are recommended.

Ancestry of the Sample: The genetic profiles of different ancestry groups are significantly different, as the allele frequencies and the correlation between genetic variants vary among different ancestry groups. The estimated effect size of a given genetic variant might be different from population to population. Therefore, the PRS generated from GWAS summary statistics using target samples is usually limited to a specific 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. Thus GWAS studies for phenotypes related to the phenotype of interest could be a substitute when GWAS studies for phenotypes of interest are not available.

Chapter 2

Polygenic Risk Score Construction

Chapter 3

Risk Prediction for Primary Ovarian Insufficiency in Childhood Cancer Survivors

Chapter 4

Conclusion

Chapter 5

Final Words

1 Lund LW, 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. *Pediatr Blood Cancer* 2011; **56**: 532–43.