Risk Prediction for Premature Ovarian Insufficiency in Childhood Cancer Survivors

by

Zhe Lu

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

The number of childhood cancer survivors has dramatically increased in the past few decades due to advances in cancer treatment, shifting the priority from clinical treatment to improving long-term survivors' quality of life. One late effect that greatly impacts female survivors is premature ovarian insufficiency (POI). It is estimated that about one in seven female survivors develops POI before age 40. POI dramatically shortens the reproductive age interval and causes infertility. To preserve the function, some fertility preservation procedures for childhood cancer survivors are now available. However, without knowing the risk of future POI, it is challenging to make fertility preservation decisions. This study aimed to build a reliable prognostic model to predict the risk of developing POI at prespecified ages in female cancer survivors to inform decision-making on fertility preservation.

We included 7,891 female survivors who are participants in the Childhood Cancer Survivor Study. The multiple imputation method was employed to deal with the missing data and an inverse probability censoring weight was assigned to each individual to account for the censoring. Elastic-Net panelized logistic regression, XGBoost, and an "Ensemble" method were used to predict the risk of experiencing POI at prespecified ages. The model performance was evaluated by nested cross-validation.

The results showed that the "Ensemble" method performed the best with AUCs (areas under the receiver operating characteristic curves) around 0.8 and AP (average positive predictive value) ranging from 0.469 to 0.595 for prespecified ages ranging from 21 to 39. The calibration curves indicated good alignment between the estimated risks of developing POI and observed status for prespecified ages less than 28. The developed "Ensemble" algorithm can be further crafted into a

user-friendly clinical tool which can provide clinicians and patients quantitative information when discussing fertility preservation.

Preface

This thesis is an original work by Zhe Lu. The research project, of which this thesis is a part, received research ethics approval from the University of Alberta Research Ethics Board, Project Name “Risk prediction for ovarian failure in childhood cancer survivors”, No. Pro00067066, August 22, 2016.

Acknowledgments

I would like to express my gratitude to all those who gave me support and assistance to complete this thesis.

I would first like to thank my supervisor, Dr. Yan Yuan, for guiding me through this research and supporting me during the graduate program. I am also grateful to Dr. Yan Yuan for providing me countless training opportunities, without that I cannot imagine completing this thesis.

I would like to thank my committee member, Dr. Paul Nathan, for taking the time to provide me helpful suggestions during the many stages of the thesis. This thesis has greatly benefited from his invaluable advice and feedback. I would like to thank Dr. Sogol Mostoufi-Moab for providing professional and valuable insight from a pediatric endocrinologist's perspective.

I would like to thank Dr. Qian Zhou for tutoring me on statistical knowledge and programming skills, and I greatly appreciate Dr. Qian Zhou's feedback on this thesis as well.

I would also like to thank Dr. Irina Dinu for taking the time to be an examiner at my thesis defense.

I would like to thank all the members of Dr. Yuan's research group, past and present, for listening to my presentations and providing valuable feedback.

Finally, I would like to thank the School of Public Health at the University of Alberta for providing me with this opportunity and support during the three years of my graduate study.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Appendix Tables

List of Appendix Figures

List of Abbreviations

AFT	Accelerated failure time
AOF	Acute ovarian failure
AP	Average positive predictive value
ASCO	American society of clinical oncology
AUC	Area under the receiver operating characteristic curve
BMT	Bone marrow transplantation
CART	Classification and regression trees
CCA	Complete-case analysis
CCS	Childhood cancer survivor
CCSS	Childhood Cancer Survivor Study
CED	Cyclophosphamide equivalent dose
CNS	Central nervous system
CV	Cross-Validation
EN-ALR	Elastic-net age-specific logistic regression
FCS	Fully conditional specification
Gy	Gray
HR	Hazard ratio
IPCW	Inverse probability-of-censoring weight
IR	Incidence rate
JM	Joint modeling
KM	Kaplan-Meier

LTFU	Loss to follow-up
MCMC	Markov chain Monte Carlo
MAR	Missing at random
MH	Menstrual health
MI	Multiple imputations
MNAR	Missing not at random
NSPM	Nonsurgical premature menopause
PH	Proportional hazards
PMM	Predictive mean matching
POI	Premature ovarian insufficiency
RF	Random forest
ROC	Receiver operating characteristic curve
RT	Radiation Therapy
sBrS	scaled Brier Score
SD	Standard deviation
SMN	Second malignant neoplasm
SPM	Surgical premature menopause
TBI	Total body irradiation
XGBoost	Extreme Gradient Boosting

1 Introduction

1.1 Background

The decline of the ovarian follicle pool in humans is a natural process. It is associated with reduced fertility in the mid-thirties, irregular menstruation from the mid-forties, and finally, follicle exhaustion and menopause in the early fifties.[1] The treatment of cancer can accelerate this process and can cause primary ovarian insufficiency (POI) which is defined as compromised gonadal function before age 40[2].

It has been shown that female childhood cancer survivors (CCSs) are at a significantly increased risk of developing POI compared to the general population.[3,4] While the prevalence of POI in the general population is about 1%[5], it was estimated that 6.3% of female CCSs lose ovarian function within 5 years of cancer diagnosis (one subcategory of POI: acute ovarian failure, AOF)[4] and 9.1% will go on experiencing nonsurgical menopause before age 40 (the other subcategory of POI: nonsurgical premature menopause, NSPM)[3].

As female CCSs face a high risk of POI which may cause many distressing chronic conditions such as infertility, osteoporosis, heart disease, and depression[6], the American Society of Clinical Oncology (ASCO) has recommended health care providers inform pediatric patients and their parents or guardians about the risk of developing POI and provide information on fertility preservation[7] such as oocyte and ovarian tissue cryopreservation[8]. However, discussing fertility preservation with pediatric patients and their families is challenging, especially when the absolute risk of developing POI in the future is unknown. Previous research has identified many cancer treatment-related risk factors of POI[3–5,9], but there is a need to build prognostic models for predicting the absolute risk of POI for individual patients.

1.2 Objective

The goal of this study was to estimate the risk of developing POI among female childhood cancer survivors before prespecified ages so that information can be provided for those survivors at high risk who might consider fertility preservation before developing POI.

1.3 Organization

The remainder of the thesis is structured as follows. In Chapter 2, I explore predictors, derived age at event, and conducted the univariate analysis. Chapter 3 addresses the analytical challenges including missing data and censoring. Chapter 4 focuses on model development and evaluation. Finally, the findings, study limitations, and future research directions are presented in Chapter 5.

2 Data

The data used to develop prognostic models was derived from the Childhood Cancer Survivor Study (CCSS), a retrospective cohort focused on the late effects of cancer treatments among CCSs. In this chapter, I first introduced the data source (sections 2.1 and 2.2) and then prepared the data including assigning age at event/censoring for each participant (section 2.3) and applying exclusion criteria (section 2.4). Finally, I conducted exploratory data analysis (sections 2.5-2.7).

2.1 Data source: The Childhood Cancer Survivor Study

The Childhood Cancer Survivor Study (CCSS) is a multi-institutional collaborative project which assembled a large and diverse cohort of childhood cancer survivors in North America, enabling investigators to study the relationship between late effects after cancer therapy and treatments.

Establishment and follow-up of the CCSS cohort

Details about the establishment and follow-up of the cohort have been documented in LL. Robison et al. (2002, 2009)[10,11] and WM. Leisenring et al. (2009)[12]. Briefly, patients were eligible for recruitment if they were 5-year survivors diagnosed before age 21 with one of the following cancer types: leukemia, central nervous system (CNS) cancers, Hodgkin lymphoma, non-Hodgkin lymphoma, neuroblastoma, soft-tissue sarcoma, kidney cancer, or bone tumors.[10] Recruitment of the original cohort began in 1992.[12,13] Investigators identified 20,720 eligible survivors diagnosed between January 1, 1970, and December 31, 1986.[11] After extensive tracing and contacting efforts, 69% (14,357) of the eligible study population completed the baseline questionnaire[12]. The demographic and cancer-related characteristics of participants, non-participants, and those who were lost to follow-up were compared in LL. Robison et al. (2002)[10] and no statistically significant

differences were found. From 2008, CCSS started to expand the original cohort by recruiting eligible survivors diagnosed between January 1, 1987, and December 31, 1999[11]. At last, the CCSS cohort was expanded to 38,036 eligible survivors with 25,665 participating[14].

To monitor the health conditions and late effects among the established cohort, five follow-up surveys had been sent out prior to the time we received the CCSS data. Among the two baseline questionnaires and five follow-up surveys[15] (shown in Figure 2.1), four questionnaires contained questions related to menstrual history (MH). The relevant questions were appended in Appendix A.

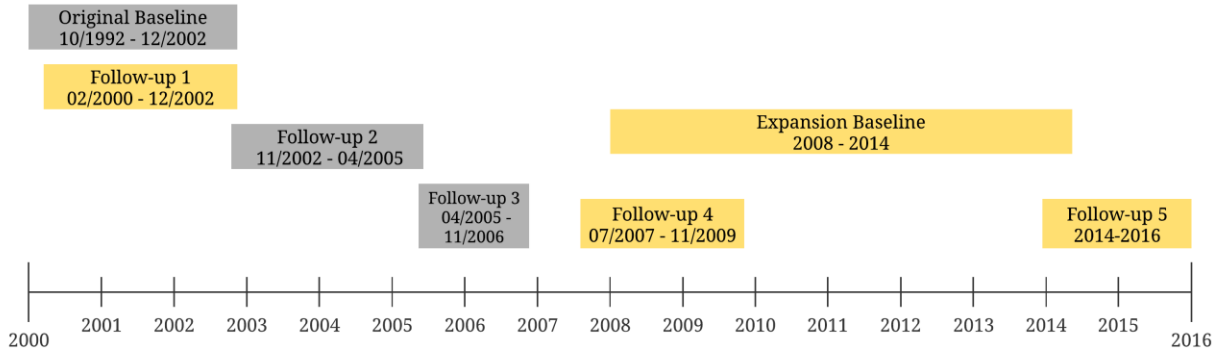


Figure 2.1 Timeline of CCSS baselines and follow-up surveys
The surveys shaded in yellow indicates that questions related to MH were included

To enable researchers to study the late effects of cancer treatments, data management staff were trained to abstract chemotherapy, radiation therapy, and surgery information from medical records by using a standardized medical records abstraction form.[12] Potential outliers were returned to the data management staff to double-check the medical records and verify data.[12]

Initial information about second malignant neoplasms (SMN) was obtained from self-reported baseline questionnaires. Then those reported positive responses were considered had possible SMNs and their information was forwarded to the CCSS Pathology Center for verification.[16] Treatment data about the SMN was not available in this study.

2.2 Relevant data elements

To develop risk prediction models for POI in CCSs, we used information from 11,336 female survivors in CCSS, including MH information (such as age at menarche, age at last menstrual period, etc) abstracted from CCSS baseline and follow-up questionnaires that were completed between November 3, 1992, and November 25, 2016[17], and information abstracted from medical charts recorded between January 1, 1970, and December 31, 2004[9]. The treatment exposure information was restricted to within 5 years of a primary cancer diagnosis. The maximum radiation doses to body regions were obtained by summing the prescribed dose from all overlapping fields in each of the respective regions. The average radiation doses to right and left ovaries were estimated separately, and the lower/higher dose was recorded as minimum/maximum ovarian radiation dose accordingly.[9] For 22 chemotherapeutic agents, the quantitative dose as well as a yes/no evaluation of exposure was abstracted. Ten widely used alkylating agents were summed to be a cumulative alkylating drug dose according to the cyclophosphamide-equivalence dose equation recorded in Green et al. (2014)[18].

2.3 Outcome variables

2.3.1 Definition of outcome

The outcome was whether a childhood cancer survivor experienced POI before a prespecified age. It was determined by the cut-off age, menstrual status, and age at the event. For example, if age 25 is the age of interest (i.e. the cut-off age), the outcome is "Yes" (or labeled as 1) when the individual experienced POI before age 25, and the outcome is "No" (or labeled as 0) if her menstrual function was normal at age 25 regardless of her menstrual status after this age. Censoring occurred when the menstrual status was normal but the age at last follow-up was below age 25, as menstrual status at age 25 was unknown. Surgical premature menopause (SPM) which refers to

premature menopause caused by surgical interventions such as bilateral oophorectomy was considered a competing risk of POI. Also, when a survivor with normal menstrual function was diagnosed with an SMN before age 25, she was deemed to be censored at age of SMN diagnosis because the treatment information of SMN was not available.

2.3.2 Ovarian status

Based on the MH information, ovarian status was assigned to CCSS participants by endocrinologists according to the established definition of ovarian status[17]. The four possible status were AOF, NSPM, SPM, and normal. AOF was assigned when a female survivor reported never experiencing menarche or permanently ceased having menses within 5 years of her cancer diagnosis[4]. NSPM was assigned when menopause, which is not related to surgical, occurred before age 40 years but after 5 years from diagnosis [3]. SPM was assigned when premature menopause was induced by hysterectomy or bilateral oophorectomy. Normal means that before age 40 or at the time of the last survey containing questions related to MH, neither of the above health conditions happened to the survivor.

In this analysis, AOF and NSPM were merged to POI which was the event of interest. SPM was the competing risk of POI.

2.3.3 Deriving age at event

Assigning age at event for individuals was not straight forward because different information, including age at menarche, age at last menstrual period, age at a most recent survey that contains MH information, age at surgical ovary removal, and age at SMN, need to be considered simultaneously. The details of deriving age at event were given in Appendix B. The algorithms for deriving it have been discussed with endocrinologists to ensure its rationality. Typically, for NSPM, age at event was from self-reported age at last menstrual period. For SPM, the age at

surgical time was used. For AOF, age at last menstrual period was used as age at event if it is available in the data set. When age at last menstrual period was missing age at menarche was used as age at event for AOF. If both ages were missing, the larger of age 16 and age at diagnosis plus 5 was used. For normal status, age at last menstrual period was used when this age was greater than 40, otherwise, when this age was less than 40, the smaller of age 40 and age at last survey that contains MH information was used as age at event.

It should be noted that 129 survivors had clear age intervals in which their ovarian function was normal, but their ovarian status became unclear after a certain age. Therefore, they were retained in the analysis until their ovarian status could not be ascertained based on the available information (the details are in Appendix B).

2.4 Exclusion criteria

To ensure the data used for analysis is accurate, survivors who did not participate in any questionnaire containing the menstrual history section (n = 1,774) and those who did not attain age 18 at their latest follow-up or provided menstrual history information through a proxy (n = 766) were excluded from the analysis. Furthermore, the subject was excluded if her menstrual status could not be determined (n = 86), or if she received $\geqslant$ 30 Gy cranial and/or pituitary radiation or was suspected pituitary dysfunction (n = 765), diagnosed as Turner or Down's Syndrome (n = 6), developed an SMN within 5 years of primary cancer (n = 21), or had conflicting age information (i.e. age at SPM, NSPM, or SMN < 9 years old and age at event occurred 3 years prior to age at diagnosis, see Appendix B) (n = 27), leaving 7,891 survivors of the original sample (70%) available for analysis (Figure 2.2). The final data set includes 7,891 observations with 61 variables. The data dictionary for the final data set was available in Appendix C.

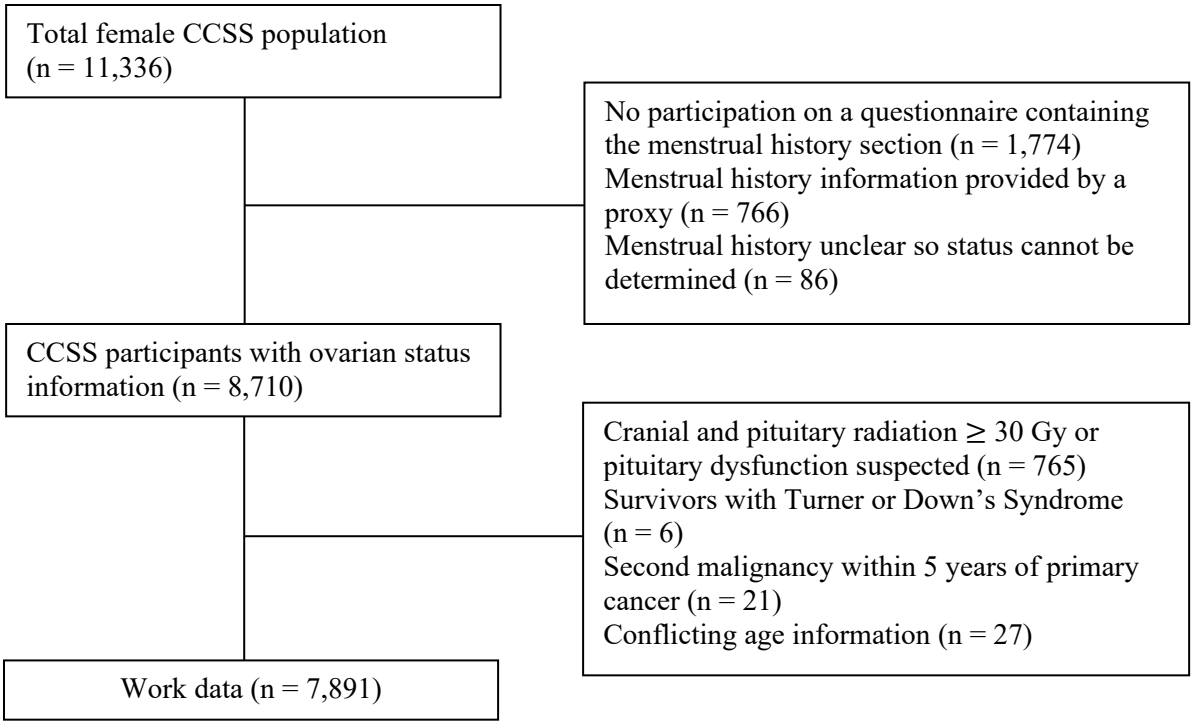


Figure 2.2 Exclusion flowchart

2.5 Exploring the elements

In the final data set, 6.2% of the total information was missing and 19% of survivors had one or more missing values. Understanding the missingness and missing pattern between variables was the first step to deal with the problem of missing. In this section, missing patterns were investigated using several visualizing methods.

Figure 2.3 showed the missing pattern of 61 variables for 7,891 observations (y-axis) in the final data set. Variables *cohort* (original or expansion), *date of birth* ("d_birth"), *ovarian status* ("status"), and *cancer type* ("diagnose") were fully observed. Among variables with missing values, *age at event* ("a_event") had the lowest missing percentage (0.48%). Variables related to cancer treatments that were extracted from medical records had higher missing percentages, ranging from 6.01% to 12.55%.

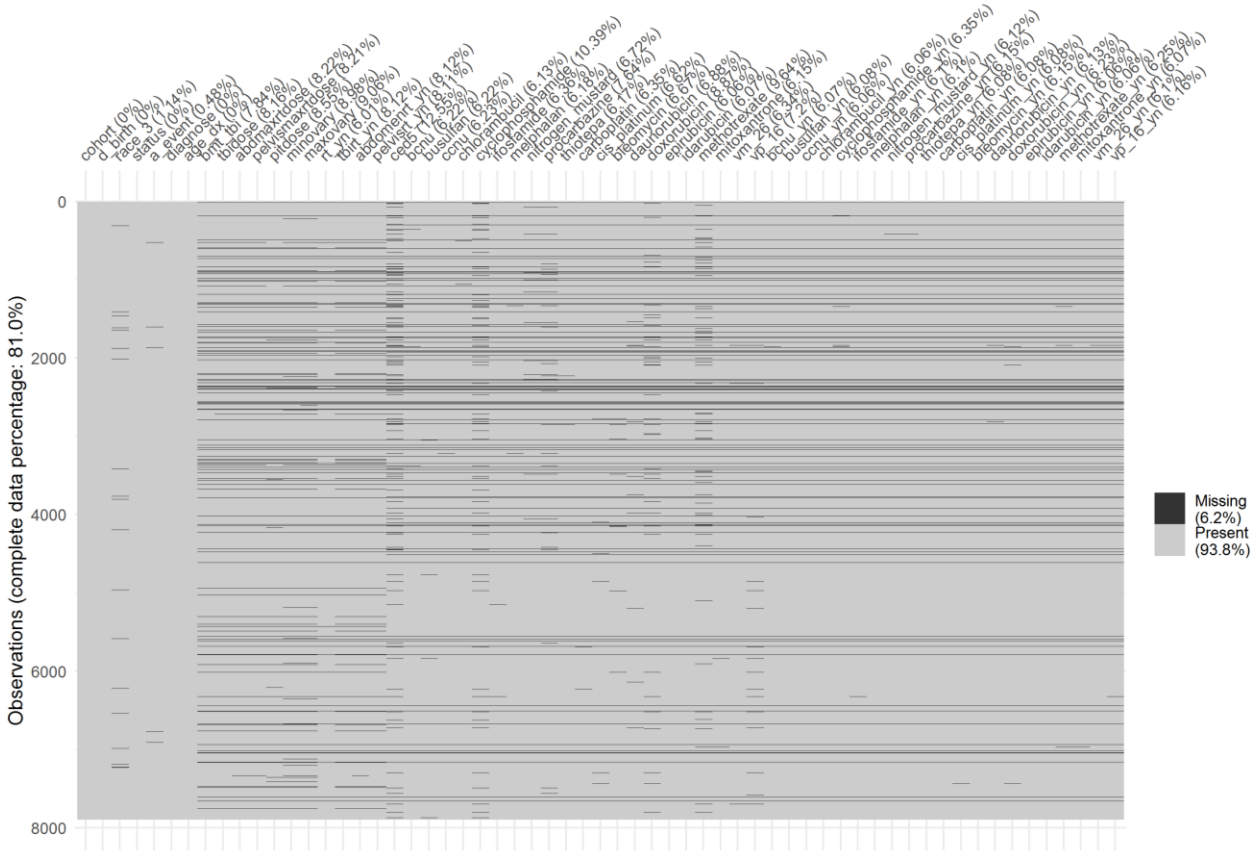


Figure 2.3 Missing pattern of final data

Figure 2.4 showed the missing percentages of variables categorized by ovarian status, the missing percentages in POI and SPM were slightly higher than that in the normal group, but the pattern among the three status was similar.

Figure 2.5 showed how the missing variables were connected. Each row represented a missing combination, and each column indicates a variable. The numbers on the left indicated the frequencies of missing with the same combinations, while the numbers on the right referred to how many variables were missing in each respective row. Finally, the bottom numbers indicated the frequencies of missing values in each variable. For example, in the first row, there were 6,319 subjects with no missing values, and in the second row, 420 subjects had missing values in 54 common variables.

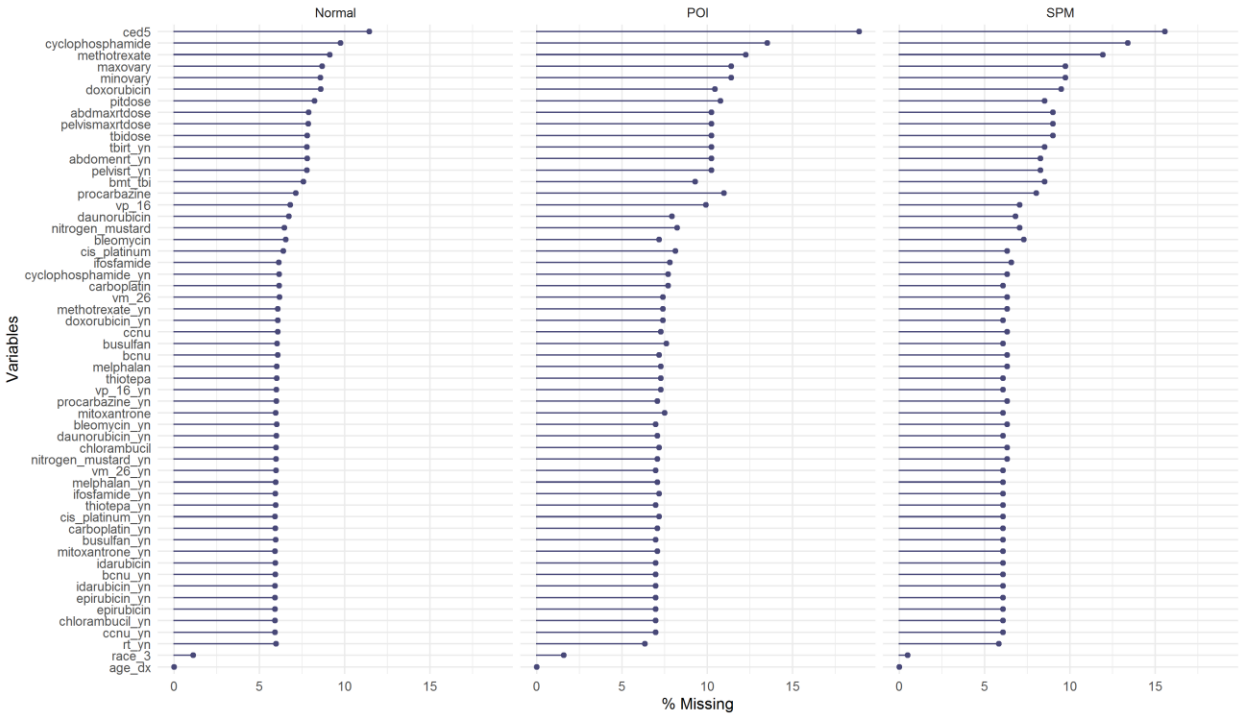


Figure 2.4 Missing percentages by ovarian status

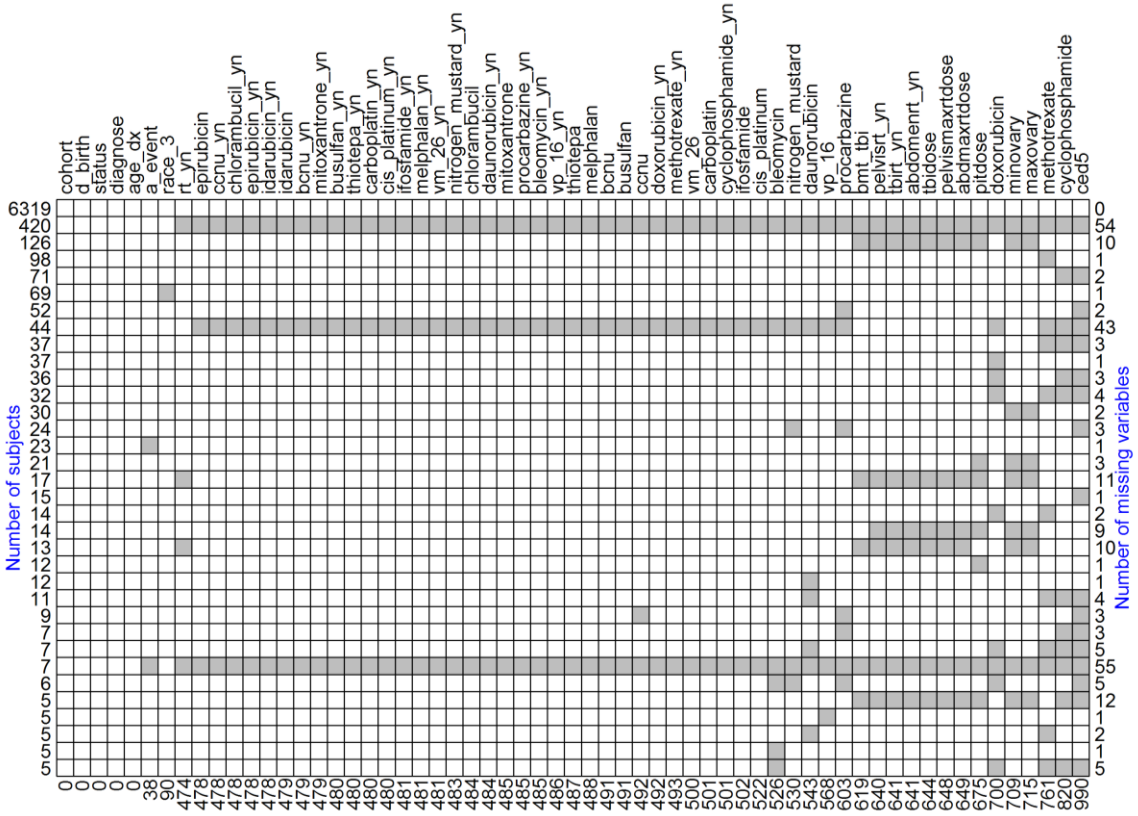


Figure 2.5 Missing patterns. (Grey boxes denote missing variables)
A partial table containing missing combinations with frequencies ≥ 5.

2.6 Exploring outcome variables

2.6.1 Distributions of age at event

Figure 2.6 showed the distribution of age at event stratified by the three levels of ovarian function. The distribution of age at POI appears skewed right. Age at SPM had a higher median age (approximately 33) compared to age at POI (approximately 18). Due to rules for assigning age at event (detailed in Appendix B), there was one peak (age = 16) in age at POI and one peak (age = 40) in age at event for survivors with normal ovarian status.

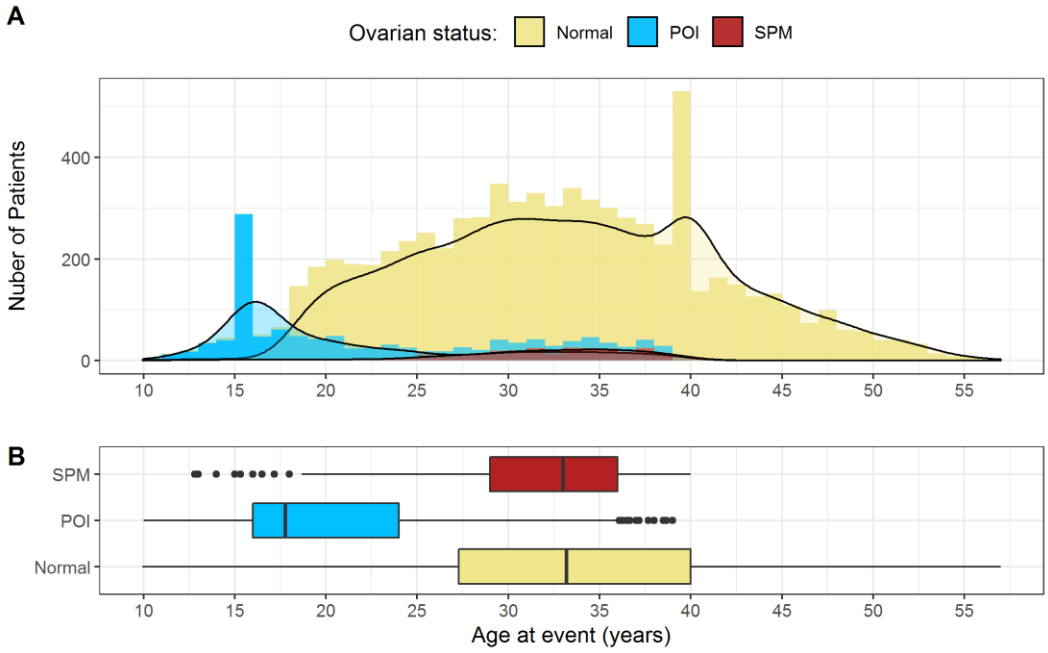


Figure 2.6 Distribution of age at event stratified by ovarian status. A) Histogram B) Boxplot.
Note: The 129 survivors with unclear status mentioned in section 2.3.3 were excluded.

2.6.2 The censoring patterns

To visualize the occurrence of censoring (the ovarian status was only partially observed before cut-off ages), Figure 2.7 showed the age at event in a different way: all the age intervals from age 9 to age at event are plotted as horizontal lines colored with three different colors. The yellow solid

line indicated that the survivor's ovarian status was normal. The blue dotted line indicated that the survivor developed POI while the red dashed line indicated that the survivor experienced SPM. As the age threshold increases, a larger proportion of survivors becomes censored (shown in Figure 2.8). The proportions of ovarian status at different age thresholds and more details on percentages of censoring were summarized in Appendix D.

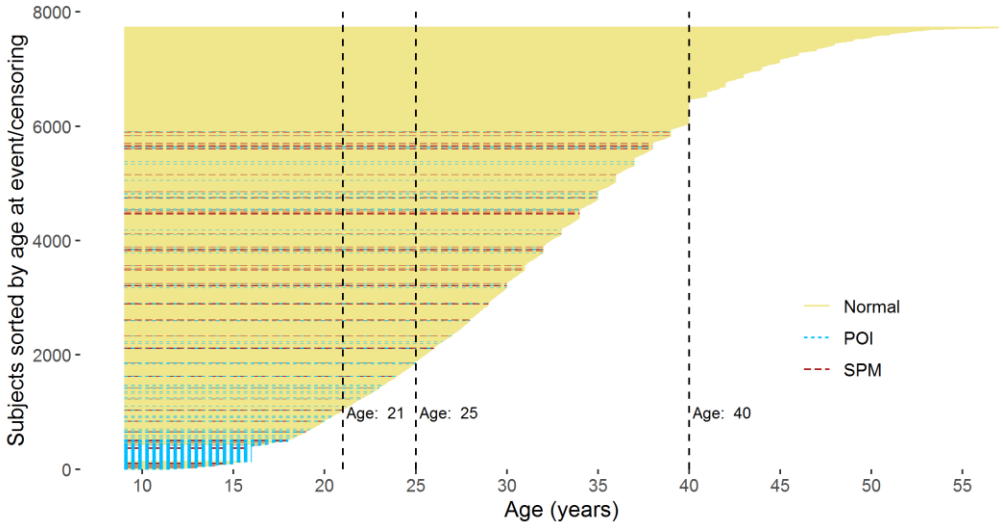


Figure 2.7 Participants sorted by their age at event or censoring.
Note: The 129 survivors with unclear status mentioned in section 2.3.3 were excluded.

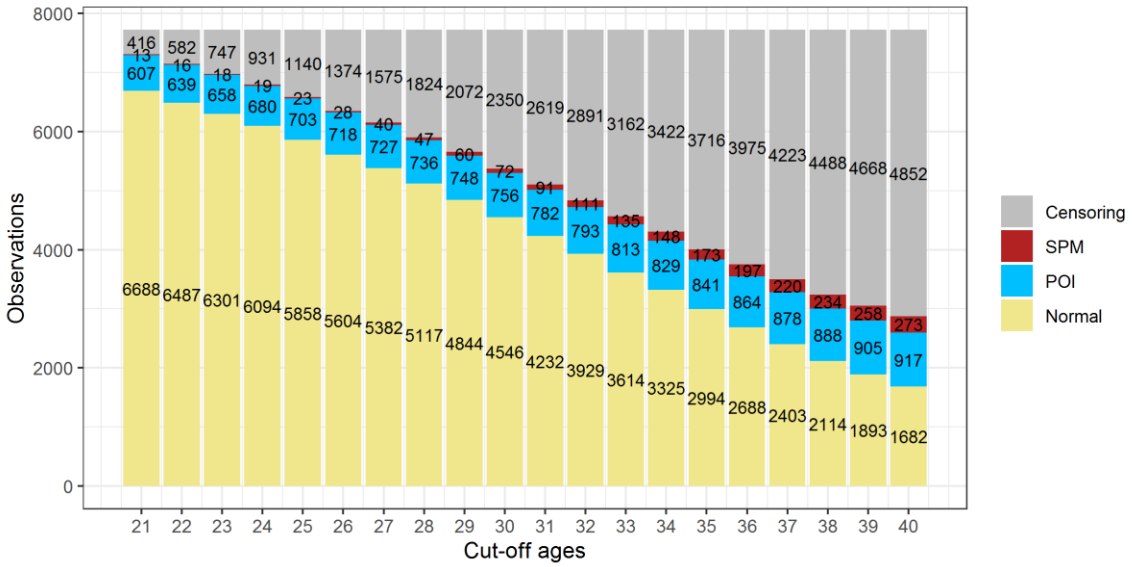


Figure 2.8 Ovarian status composition in CCSS participants at different age thresholds
Note: The 129 survivors with unclear status mentioned in section 2.3.3 were excluded.

2.6.3 Cumulative Incidence Curves

Figure 2.9 showed the cumulative incidence curve of POI and SPM respectively. A large proportion of AOF (one type of POI) experienced the event before age 20; therefore, the cumulative incidence of POI increased rapidly at the beginning, especially at age 16 where there was a big jump due to the rule of assigning age at event. After 25, the cumulative incidence of POI went up steadily and reached about 15% by age 40. This number was close to the sum of the prevalence of AOF (6.3%) and cumulative incidence of NSPM (8%) reported in the previous CCSS reports[3,4,19]. The cumulative incidence of SPM increased slowly at the beginning but accelerated more rapidly after surpassing 25 years old until reaching around 7% at age 40.

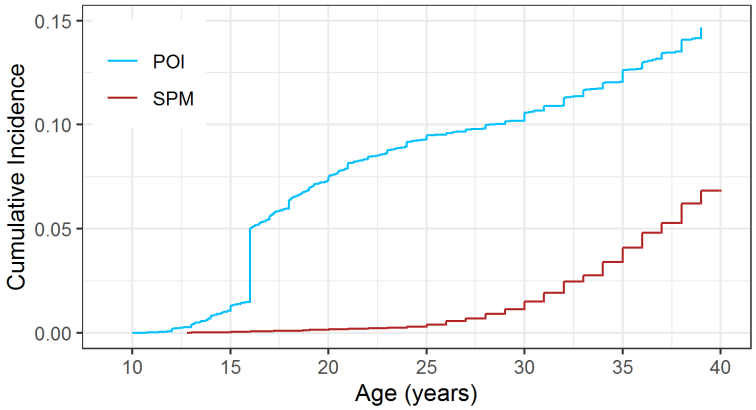


Figure 2.9 Cumulative incidence of POI and SPM

2.7 Predictors

Risk factors for developing POI have been identified in the literature. They were irradiation to the abdomen or pelvis, total body irradiation, and alkylating chemotherapy agents.[4,20–22] Other than cancer treatments, age at diagnosis (or age at the time of treatment) and BMT were also regarded as related to the risks of developing POI.[23]

In this section, I inspected all the available exposure variables in the final data set and explored their relationships with an age-specific ovarian status (age threshold = 25). Table 2.1 showed the distribution of the age-specific ovarian status.

Table 2.1 Distribution of the ovarian status when age threshold = 25

Cut-off age (years)	Normal n (%)	SPM n (%)	POI n (%)	Censoring n (%)	Total
25	5968 (75.8)	23 (0.3)	703 (8.9)	1140 (14.5)	7834[a]

a: *19 of the 129 survivors (section 2.3.3) were excluded because their menstrual status was unclear before 25 years; 38 subjects whose age at event were missing were also excluded.*

2.7.1 Race

The self-reported race was merged into 3 categories: *white*, *black*, and *other*. The percent of self-reported *white* was 86.1%, while the percentages of *black* and *other* races were 6.1% and 6.6%, respectively. The missing proportion was 1.1%.

Table 2.2 showed the distribution of race. 8.6% of *white* developed POI, while higher percentages of POI were seen in *black* (9.8%) and *other* (12.0%), which indicated that race might be a risk factor and *black* and *other* had a higher risk of developing POI comparing to white people.

Table 2.2 Distribution of races by ovarian status at age 25 (%: row percent)

Race	Normal n (%)	POI n (%)	SPM n (%)	Censored n (%)	Overall
White	5221 (77.3)	582 (8.6)	21 (0.3)	926 (13.7)	6750
Black	335 (70.0)	47 (9.8)	1 (0.2)	95 (19.9)	478
Other	356 (68.7)	62 (12.0)	1 (0.2)	99 (19.1)	518
Missing	56 (63.6)	12 (13.6)	0 (0)	20 (22.7)	88
Total	5968 (76.2)	703 (9.0)	23 (0.3)	1140 (14.6)	7834

2.7.2 Age at diagnosis

Figure 2.10 showed the distribution of age at diagnosis stratified by ovarian status by age 25. Visually, there were two peaks at approximately 4 years old and 14 years old for Normal and POI. SPM had only 23 observations (Table 2.1), therefore was omitted in Figure 2.10A.

The boxplots in Figure 2.10 showed that the median age at diagnosis was slightly younger in POI compared to Normal and SPM. However, this should not lead to the conclusion that patients diagnosed at a younger age tend to develop POI, because the boxplots did not take censoring information into account, and age at diagnosis was strongly associated with censoring (Figure 2.10A) as patients diagnosed at a younger age had more potential to be younger than age 25 at the time of the last follow-up (for example, if a patient diagnosed at age 2 years old in 1999, she only attained 20 in 2018).

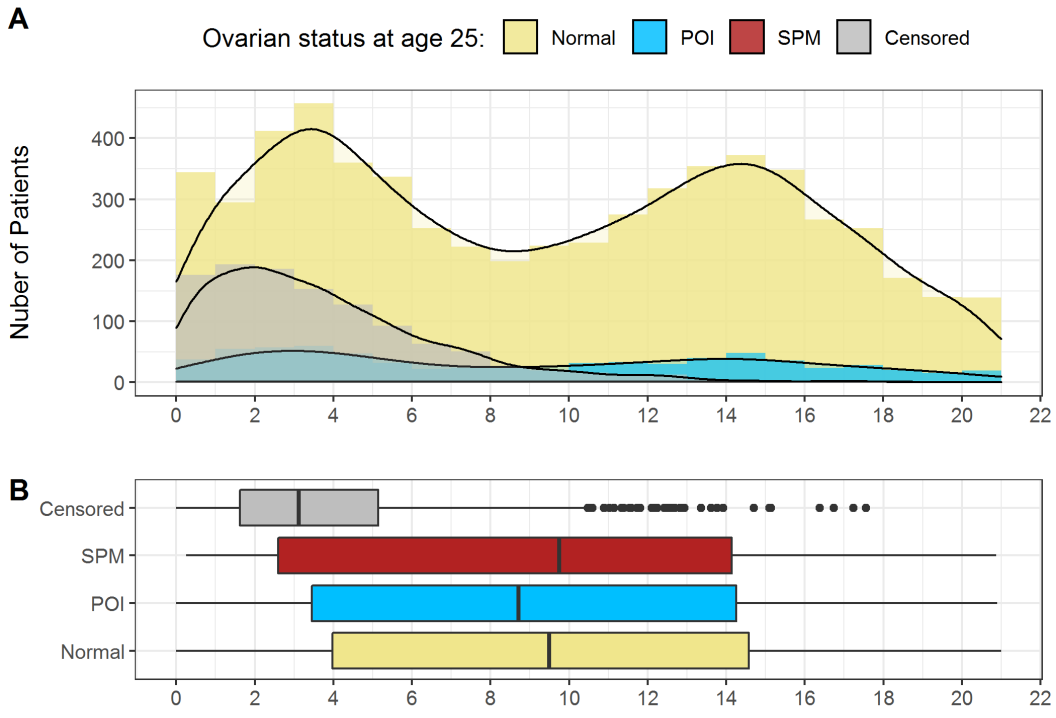


Figure 2.10 Distribution of age at diagnosis
A) Histogram and density curve; B) Boxplots

2.7.3 Radiotherapy

Irradiation can cause ovarian damage, and the degree of the radiation-induced damage depends on the radiation dose, radiation field, fractionation schedule, and age at the time of treatment.[24–26] Research has shown that irradiation to the abdomen and pelvis appears to lead to the highest risk of developing POI.[4,24,26,27] As for the irradiation doses, it has been reported that receiving pelvic or abdominal radiation doses ≥10 Gy for post-pubertal girls and ≥15 Gy for prepubertal girls substantially increased the risk of POI.[28]

Table 2.3 showed that a similar proportion (around 44%) of patients in the study sample received irradiation to the abdomen, pelvis, pituitary, and both ovaries.

Table 2.3 Radiotherapy Information (%: row percentages, age cut off 25)

Irradiation	Not received n (%)	Received n (%)	Missing n (%)	Total n
Abdomen	3707 (47.3)	3487 (44.5)	640 (8.2)	7834
Pelvis	3707 (47.3)	3488 (44.5)	639 (8.2)	
Pituitary	3721 (47.5)	3446 (44.0)	667 (8.5)	
Ovary (minimum)[a]	3708 (47.3)	3426 (43.7)	700 (8.9)	
Ovary (maximum)[b]	3708 (47.3)	3420 (43.7)	706 (9.0)	

a: the minimum of the estimated average dose between right and left ovaries.
b: the maximum of the estimated average dose between right and left ovaries.

Figure 2.11 showed the density curves of radiation doses (only doses > 0) to different body regions categorized by ovarian status at age 25 (density curves for SPM were not plotted as SPM has only 23 observations). It should be noted that the patients who received more than 30Gy irradiation to pituitary have been excluded. Visually, irradiation dose to abdomen, pelvis, and both ovarian doses could well separate Normal and POI groups, indicating that they might be useful risk factors in predicting the risk of developing POI before 25 years old.

It should be noted that the concern of bias caused by the censored information, however, was not a big issue in analyzing the effect of treatment exposures as these treatments were unlikely to be associated with censoring.

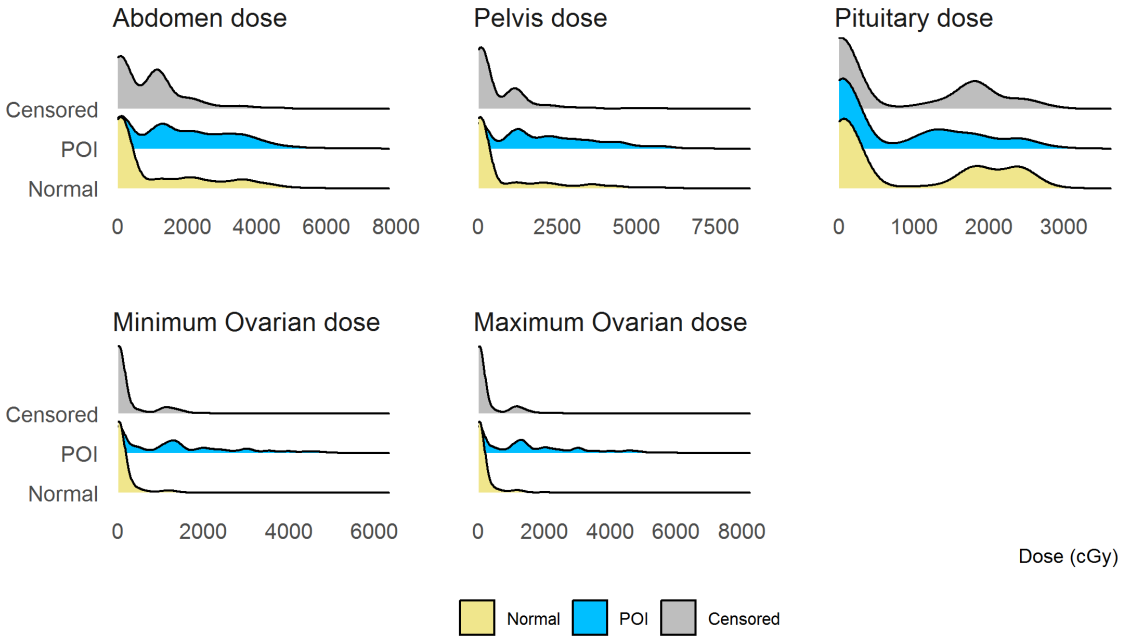


Figure 2.11 Distribution of radiation doses (> 0) categorized by ovarian status at 25.

2.7.4 Chemotherapy

Research has shown that gonadotoxic chemotherapy can bring ovaries damage ranging from low (can be recovered to normal ovarian function) to serious (ovarian atrophy and complete ovarian failure)[20]. The degree of chemotherapy-induced damage depends on agents' type and dose[27].

Alkylating agents are widely used for the treatment of pediatric cancers – these agents inhibit cancer cell division by breaking DNA strands. Therefore, they are also toxic to normal cells, particularly those sensitive to DNA damage, which includes oocyte and follicular cells in ovaries.[29] The impact of alkylating agents on ovaries has been assessed by histological analysis.[30] Alkylating

agents result in significantly lower counts of primordial follicles compared with those who have not received chemotherapy and even those who received non-alkylating agents.[30] The identified high risk alkylating agents include Cyclophosphamide, Busulfan, Melphalan, Procarbazine, and Ifosfamide[23]. To estimate cumulative alkylating agent exposure's effect, the Cyclophosphamide Equivalent Dose (CED) was developed to allow the comparison of different common alkylating agents.[18]

Doxorubicin, an anthracycline (a type of anti-tumor antibiotics) was used to treat a wide range of cancers. Similar to alkylating agents, these agents can also cause DNA double-strand breaks leading to the death of primordial follicles.[31] Recent studies have suggested doxorubicin is most closely linked to ovarian failure among the non-alkylating agents.[20,32–34]

Cisplatin is a heavy metal-based compound and acts as a DNA cross-linking agent that interferes with DNA repair mechanisms, blocking cell division.[35] Its mechanisms of ovarian toxicity has not been well studied because female patients treated with cisplatin have usually received it as part of a multiple-drug regimen.[32] According to recent studies, it was considered as a moderate risk factor for ovarian failure[35,36].

VP-16 (etoposide), a topoisomerase inhibitor, unwinds during DNA replication and break DNA double-strand like cisplatin. Therefore, VP-16 is especially toxic to cells that are sensitive to DNA damage, such as granulosa cells and oocytes.[37]

Table 2.4. summarized the risk of ovarian toxicity of chemotherapy agents (which were available in our data set) reported in previous research.

Table 2.5 showed the proportion of patients that received each type of chemotherapy. The three most used chemotherapy agents (shaded in yellow) were cyclophosphamide, methotrexate, and doxorubicin received by 2890 (36.7%), 2789 (35.4%), and 2647 (33.6%) patients, respectively. Other chemotherapy agents were much less used in this study sample. For example, eight chemotherapy agents (shaded in blue) were used by less than 100 patients in this study sample.

Table 2.4 Risk of ovarian toxicity of chemotherapy agents in literature

Type	Chemotherapy agent name	Risk	Reference
Alkylating agent[38]	BCNU (Carmustine)	High risk	39
	Busulfan	High risk	4,22,23,40
	CCNU (Lomustine)	High risk	4
	Chlorambucil	High risk	4,22,40
	Cyclophosphamide	High risk	4,22,23,40
	Ifosfamide	High risk	23
	Melphalan	High risk	22,23
	Nitrogen mustard	High risk	4,22
	Procarbazine	High risk	4,22,23,40
	Thiotepa	High risk	39
Platinum compounds[41]	Carboplatin	Intermediate risk	23
	Cisplatin	Intermediate risk	22,23,40
Antibiotics	Bleomycin	Low risk	22,23,42
Anthracyclines, anthraquinone[38]	Daunorubicin	NA	
	Doxorubicin	Intermediate risk	40
	Epirubicin	NA	
	Idarubicin	NA	
	Mitoxantrone	NA	
Antimetabolites[38]	Methotrexate	Low risk	22,23,42
Epipodophyllotoxins[38]	VM-26 (Teniposide)	NA	
	VP-16 (Etoposide)	Low risk	23

Table 2.5 Chemotherapy Information

Type	Chemotherapy agent	Not Received n (%)	Received n (%)	Missing n (%)	Overall n
Alkylating agent	BCNU	7214 (92.1)	137 (1.7)	483 (6.2)	7834
	Busulfan	7290 (93.1)	61 (0.8)	483 (6.2)	
	Cyclophosphamide	7288 (93)	62 (0.8)	484 (6.2)	
	Chlorambucil	7342 (93.7)	16 (0.2)	476 (6.1)	
	CCNU	4149 (53)	2876 (36.7)	809 (10.3)	
	Ifosfamide	7005 (89.4)	336 (4.3)	493 (6.3)	
	Melphalan	7272 (92.8)	82 (1)	480 (6.1)	
	Nitrogen Mustard	6990 (89.2)	324 (4.1)	520 (6.6)	
	Procarbazine	6638 (84.7)	604 (7.7)	592 (7.6)	
	Thiotepa	7328 (93.5)	27 (0.3)	479 (6.1)	
P[a]	Carboplatin	7213 (92.1)	128 (1.6)	493 (6.3)	
	Cis_Platinum	6874 (87.7)	446 (5.7)	514 (6.6)	
A[b]	Bleomycin	6758 (86.3)	558 (7.1)	518 (6.6)	
Anthracyclines, anthraquinone	Daunorubicin	6180 (78.9)	1119 (14.3)	535 (6.8)	
	Doxorubicin	4508 (57.5)	2635 (33.6)	691 (8.8)	
	Epirubicin	7364 (94)	0 (0)	470 (6)	
	Idarubicin	7305 (93.2)	58 (0.7)	471 (6)	
	Mitoxantrone	7328 (93.5)	29 (0.4)	477 (6.1)	
A[c]	Methotrexate	4306 (55)	2778 (35.5)	750 (9.6)	
E[d]	VM 26	7052 (90)	290 (3.7)	492 (6.3)	
	VP 16	6319 (80.7)	957 (12.2)	558 (7.1)	

a) P: Platinum compounds

b) A1: Antibiotics

c) A2: Antimetabolites

d) E: Epipodophyllotoxins

Figure 2.12 showed the density curves of chemotherapy agents' doses (only doses > 0) categorized by ovarian status at age 25 (density curves for SPM were not plotted as SPM has only 23 observations) for chemotherapy agents' doses. The top two rows were ten alkylating agents, among which, busulfan, chlorambucil, melphalan, and thiotepa appeared to be good markers that can separate POI and normal. Other alkylating agents however had similar distribution between POI and normal. Among non-alkylating chemotherapy agents, the dose of mitoxantrone seemed to be a good marker for distinguishing POI and Normal. Doxorubicin, which was identified as closely linked to ovarian failure in the literature[40], however, seemed to have similar distribution between POI and Normal.

It should be noted that the density curves for chemotherapy agents that were not widely used cannot be reliably estimated, and the plots in Figure 2.12 were marginal analysis which did not consider the effect of confounders.

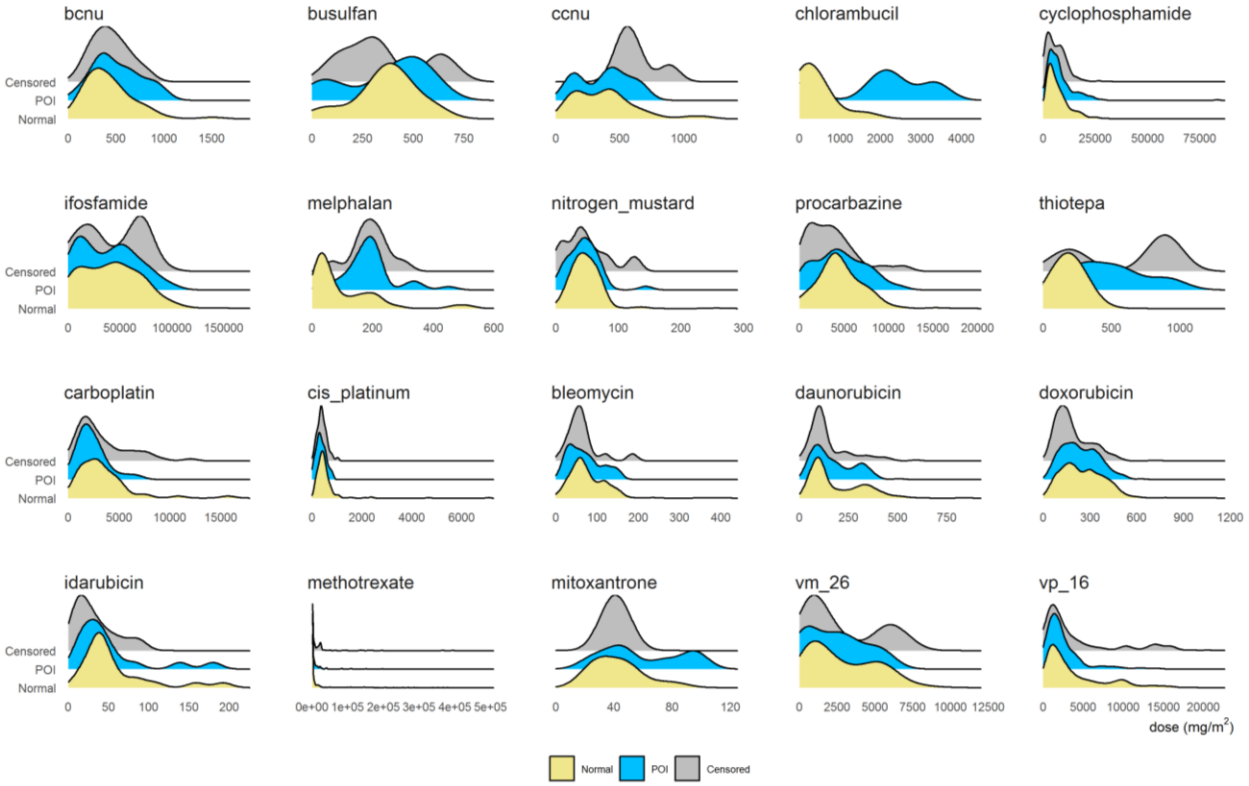


Figure 2.12 Distribution of Chemotherapy dose (>0) categorized by ovarian status at age 25

2.7.5 BMT

Table 2.6 showed the percentages of BMT categorized by ovarian status at age 25. A higher proportion (50.4%) of survivors who experienced BMT developed POI comparing to those who did not experience BMT (6.9%), suggesting BMT was an important risk factor for POI.

Table 2.6 BMT categorized by ovarian status at age 25 (%: row percentages)

BMT	Normal n (%)	POI n (%)	SPM n (%)	Censored n (%)	Overall
No	5371 (77.6)	475 (6.9)	22 (0.3)	1023 (14.8)	6891
Yes	113 (33.5)	170 (50.4)	0 (0)	50 (14.8)	333
Missing	484 (78.3)	58 (9.4)	1 (0.2)	67 (10.8)	610
Total	5968 (76.2)	703 (9.0)	23 (0.3)	1140 (14.6)	7834

Total body irradiation (TBI), one of the conditioning regimens given before bone marrow transplantation, was identified as an important risk factor for premature ovarian failure[43].

Table 2.7 showed the frequencies of patients who received TBI categorized by ovarian status at age 25. A higher proportion (59.4%) of survivors who received TBI developed POI compared to those who did not receive TBI (7.7%), suggesting BMT was an important risk factor of POI.

Table 2.7 TBI categorized by ovarian status at age 25 (%: row percentages)

TBI	Normal n (%)	POI n (%)	SPM n (%)	Censored n (%)	Overall
Not Received	5427 (77.1)	541 (7.7)	22 (0.3)	1049 (14.9)	7039
Received	48 (30.0)	95 (59.4)	0 (0.0)	17 (10.6)	160
Missing	493 (77.6)	67 (10.6)	1 (0.2)	74 (11.7)	635
Total	5968 (76.2)	703 (9.0)	23 (0.3)	1140 (14.6)	7834

Figure 2.13 showed the density curves for TBI doses (only doses > 0) categorized by ovarian status at age 25 (density curves for SPM were not plotted as SPM has only 23 observations). Visually, the three curves had similar distributions, although the peak in POI was at a slightly higher dose compared to the peak in Normal.

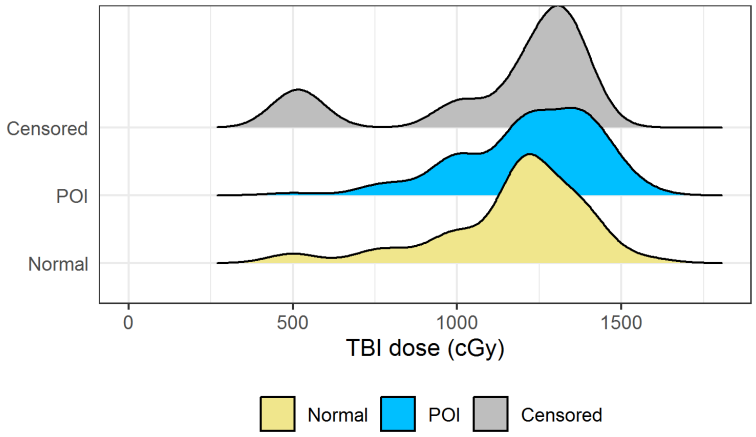


Figure 2.13 Density curve of TBI dose (> 0) categorized by ovarian status at age 25

2.7.6 Diagnosis

Table 2.8 showed the distribution of eight cancer diagnosis categories by ovarian status at age 25. Compared to leukemia which had a similar POI proportion to the overall POI proportion, neuroblastoma, kidney tumors, Hodgkin lymphoma, and soft tissue sarcoma had a higher proportion of POI, while bone cancer, CNS, and non-Hodgkin lymphoma had a lower proportion of POI.

Figure 2.14 showed the treatment patterns in different cancer diagnosis types. Cancer diagnoses contained information about treatment regimens, and therefore it may be useful to help improve the accuracy of predicting POI.

Table 2.8 Distribution of eight cancer diagnosis categories by ovarian status at age 25 (%: row percentages)

Diagnose	Normal n (%)	POI n (%)	SPM n (%)	Censored n (%)	Overall
Leukemia	1964 (76.3)	234 (9.1)	4 (0.2)	370 (14.4)	2572
Neuroblastoma	391 (61.1)	63 (9.8)	1 (0.2)	185 (28.9)	640
Kidney tumors	593 (63.5)	91 (9.7)	3 (0.3)	247 (26.4)	934
Hodgkin Lymphoma	1035 (87.9)	118 (10.0)	1 (0.1)	23 (2.0)	1177
Soft tissue sarcoma	351 (78.2)	59 (13.6)	8 (1.8)	31 (6.9)	449
Bone cancer	685 (88.2)	42 (5.4)	2 (0.3)	48 (6.2)	777
CNS	530 (69.5)	61 (8.0)	2 (0.3)	170 (22.3)	763
Non-Hodgkin lymphoma	419 (80.2)	35 (6.7)	2 (0.4)	66 (12.6)	522
Total	5968 (76.2)	703 (9.0)	23 (0.3)	1140 (14.6)	7834

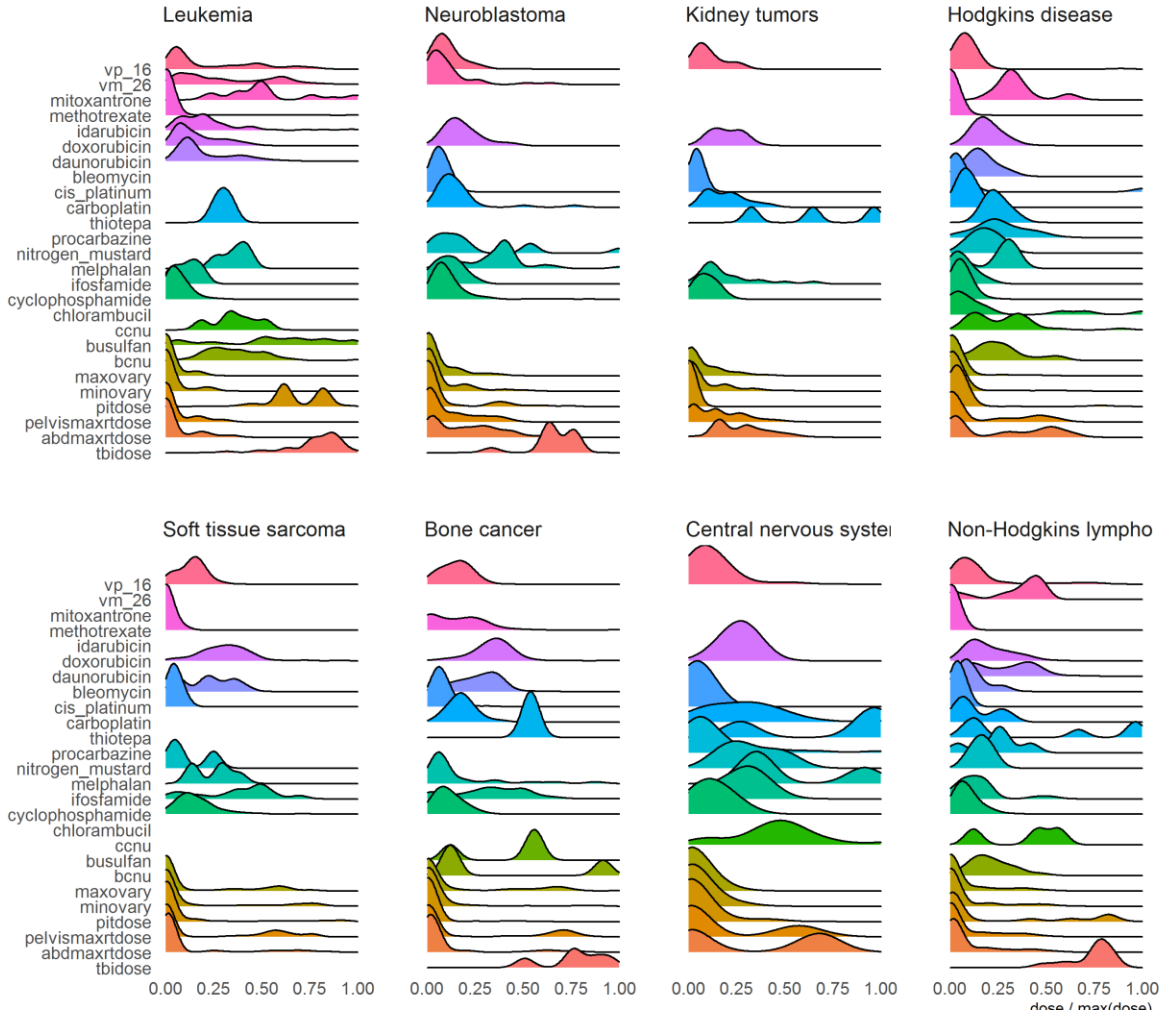


Figure 2.14 Treatment combinations in different cancer diagnosis

3 Analytical challenges and approaches

This chapter addressed challenges related to data issues including missing data and censoring. About 19% of observations had one or more missing values. Ignoring them or inappropriately dealing with them may lead to a biased inference. Section 3.1 introduced the method of multiple imputation (MI) and the procedure for conducting MI. Also, this section addressed the conflict between MI and validation methods to avoid over-optimism in the evaluation of model performance. For censoring, (i.e. some survivors with normal ovarian status did not reach the age threshold at the time they filled out the most recent survey containing MH information), section 3.2 employed the idea of inverse probability weights and highlighted how this method derived unbiased inference and dealt with the competing risk SPM.

3.1 Multiple Imputation

3.1.1 Introduction

Typically, missing data is dealt with by using the "complete-case analysis" (CCA) which ignores incomplete observations and runs analysis on the complete ones. Although this method is easy to implement and used widely in the public health field, it uses the data inefficiently because an observation will be excluded even if it has a missing value in only one variable. For example, in this project, only 6.3% of the total data was missing; however, by using CCA, almost one in five observations would be removed. Other than inefficiency, CCA may lead to biased inference when the missingness is associated with response variables.[44,45]

To improve the inefficiency and reduce bias, MI was proposed by Rubin in 1978[46] and developed in the following decades [47–49]. MI refers to the procedure of replacing each missing value with

multiple imputed values,[44] which can be thought of as drawing multiple values from the distribution of the variables given the observed data.[44] Therefore MI returns multiple complete data sets, thus ensuring all the observed data are used to develop models. It has been theoretically proved that MI can produce valid inference in terms of unbiased parameter estimates and unbiased variance estimates under the assumption "missing at random" (MAR). MAR means missingness only depends on the observed data.[44]

In contrast to MAR, "missing not at random" (MNAR) means that the missingness of the variable is, however, associated with the unobserved value of this variable.

Distinguishing the missingness mechanism is at the core of choosing an appropriate imputation method. However, it is impossible to differentiate MAR from MNAR based on observed data alone, because the assessment requires us to know the value of missing data.

In practice, we must use our knowledge about the study to decide whether MAR is plausible. According to the analysis in Chapter 2, we know that the missing data almost always happened in treatment variables that were abstracted from medical records. Such missingness was likely to be random when the information was collected from different hospitals rather than depending on the value of these treatments. Therefore, we can assume MAR is plausible in this study.

3.1.2 Implementation of MI

When more than one variable has missing values, we need to conduct multivariate imputation which is more challenging compared to univariate imputation. There are two general approaches for multivariate imputation: joint modeling (JM)[50] and fully conditional specification (FCS)[51]. While JM requires an assumption about a common prior multivariate distribution (often specified

as the multivariate normal distribution) for all the variables, FCS relaxes this assumption by specifying the imputation model for each variable.

In this project, we employed the FCS approach in R package "mice"[52,53]. It conducts Multivariate Imputation by Chained Equations. In this algorithm, each variable with missing data is modeled in turn conditional upon the other variables in the data. Suppose there are p variables that have missing values and M is the number of iterations, the steps of MI can be summarized as:

Step 1: Specify an imputation model for variable X_j with $j = 1, \dots, p$;

Step 2: For each j, fill in initial imputations $\dot{X}_j^0$ by random draws from X_j^{obs}.

Step 3: Repeat for $t = 1, \dots, M$

Step 4: Repeat for $j = 1, \dots, p$

Step 5: Define $\dot{X}_{-j}^t = \left(\dot{X}_1^t, \dots, \dot{X}_{j-1}^t, \dot{X}_{j+1}^t, \dots, \dot{X}_p^t\right)$ as the current complete data except X_j

Step 6: Build models on data $\left(X_j^{obs} \mid \dot{X}_{-j}^t\right)$

Step 7: Draw imputations from predicted $X_j^{missing}$ from models built in step 6

Step 8: End repeat j

Step 9: End repeat t,

where X_j^{obs} means the observed values in X_j and $\dot{X}_j^t$ means imputed X_j after tth imputation.

The number of multiply imputed data sets: **m**

Many researchers have investigated the influence of **m** on different aspects of results.[54–58] Basically, higher **m** brings benefits in terms of smaller standard errors and more replicable results. However, when the primary interest is only on point estimates, (for example, risk prediction) the

high **m** may not be worth the time the procedure takes. In this project, we set **m** to be 5 considering the computation time.

3.1.3 Combining MI and nested CV

Cross-Validation (CV) is an internal validation technique, which randomly divides data set into k folds (groups) and then develops models on k–1 folds leaving out one-fold for validation. This procedure was repeated k times to utilize every fold as a validation set once. Nested CV[59] goes one step further by constructing an inner CV in k–1 folds and utilizing the remaining fold as a test set similar to an external test set (details available in Appendix E).

An honest evaluation requires that the validation set be independent of the training set. However, the MI process uses all the observed data to predict the missing values, which "leaks" the information from the validation sets to training sets. To prevent this situation, we removed the observed values of outcome variables (menstrual status and age at event) of survivors with missing covariates in validation sets (including inner and outer validation sets in nested CV, see Appendix E) before imputation and put them back after the completion of imputation in validation sets. Figure 3.1 illustrated the procedure of combining CV and MI.

It should be noted that every time the validation set switched, different outcome variables need to be removed accordingly, which required a new set of multiple imputations. Therefore, combining multiple imputation (m) and nested CV (k-fold outer CV and l-fold inner CV) will generate $m \times k \times l$ complete data sets. In this study, $m,\ k, and\ l$ were all set to be 5, thus there were 125 data sets generated after the MI process. They were all used for calculating IPCW and modeling the risk of POI at different ages.

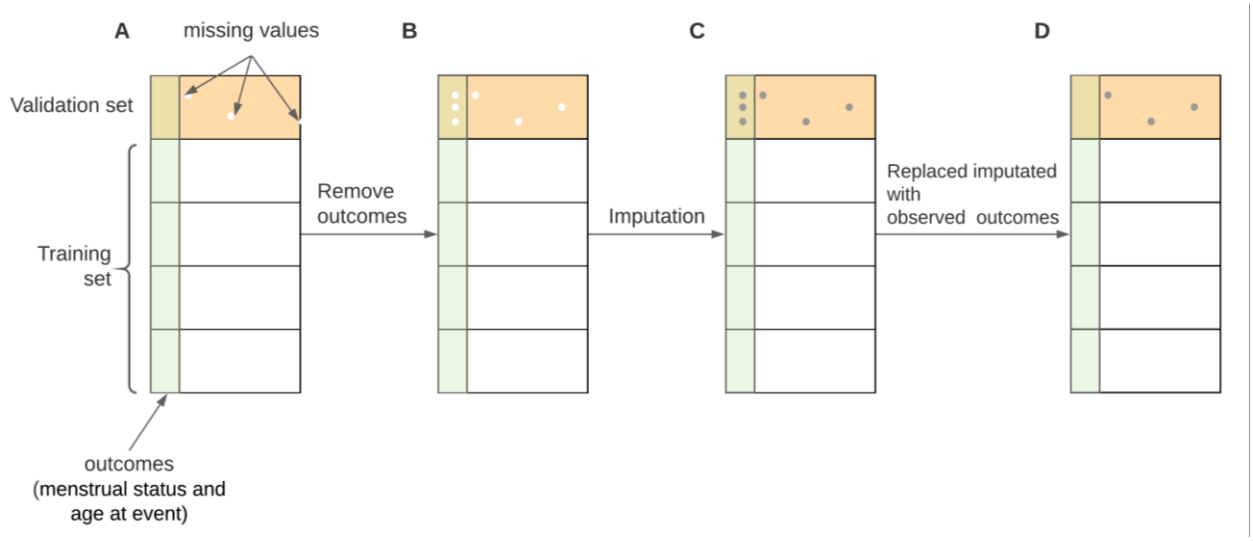


Figure 3.1 Procedure for combining MI and CV.
From A to B, the outcome variables of survivors with missing covariates in validation sets were removed; from B to C, multiple imputation was conducted for both missing covariates and removed outcome variables, from C to D, original outcome variables were put back to their positions.

3.1.4 Imputation model

In the "mice" package, various imputation methods have been built in to predict the missing values. For continuous variables, unconditional mean imputation, Bayesian linear regression, linear regression using bootstrap, etc. are available. For categorical variables, we can choose logistic regression, proportional odds models, or polytomous logistic regression. Also, predictive mean matching (PMM), classification and regression trees (CART), and random forest (RF) can be used for both continuous and categorical variables.

Among these methods, PMM avoids implausible values (e.g. negative doses) and takes heteroscedastic data into account more appropriately due to its algorithm design. PMM identifies a small subset of observations (typically up to 10) that have similar values to the predicted value for the missing entry and then drawn randomly from these candidates.[53] Besides, this procedure is

much faster than another robust non-parametric algorithm RF. In this study, PMM was employed to impute all the continuous and categorical variables.

3.1.5 Convergence and the iteration number

Since the imputation was implemented using MCMC, the iteration number M should be large enough to allow the imputations to attain convergence. Appendix F assessed the convergence of imputation process. It turns out that convergence was attained when iteration number ≥ 10. Therefore, the iteration M in this study was set to be 10.

3.1.6 Post-processing

In this project, age at event was missing in 38 subjects. The missing age at event relates to two ovarian status' levels: SPM and NSPM. Based on the definition, age at NSPM should fall in the interval between (age at diagnosis + 5) and 40 (not included), and age at SPM should be between age at diagnosis and 40. To ensure the imputed ages at the event are within reasonable ranges, "post-process" functions provided in the 'mice' package were employed. This method only affected the synthetic values and left the observed data untouched.

The value of some variables implied the value of another variable. For example, if the dose of the irradiation to a certain body region (or a chemotherapy agent) is positive, the variable indicating patients received this irradiation (or the chemotherapy agent) should be "Yes"; otherwise, it should be "No". For these situations, doses were imputed firstly, and the missing "Yes/No" indicators were assigned accordingly after dose imputation.

3.1.7 Results

All the missing values were successfully imputed, i.e., 5 imputation values for each missing data were generated. Because there were 25 different validation sets, a total of 125 complete data sets were obtained.

Figure 3.2 and Figure 3.3 showed the smoothed density estimation of original and imputed data for irradiation doses and chemotherapy agents' doses, respectively. They only showed the positive doses and those with less than 20 positive doses were omitted. The imputed data and observed data had similar distributions.

Table 3.1 summarized the numbers and percentages of missing and positive values for the irradiation doses and chemotherapy agents' doses in the observed data set and the imputed data. The imputed data had similar percentages of positive value with the observed data.

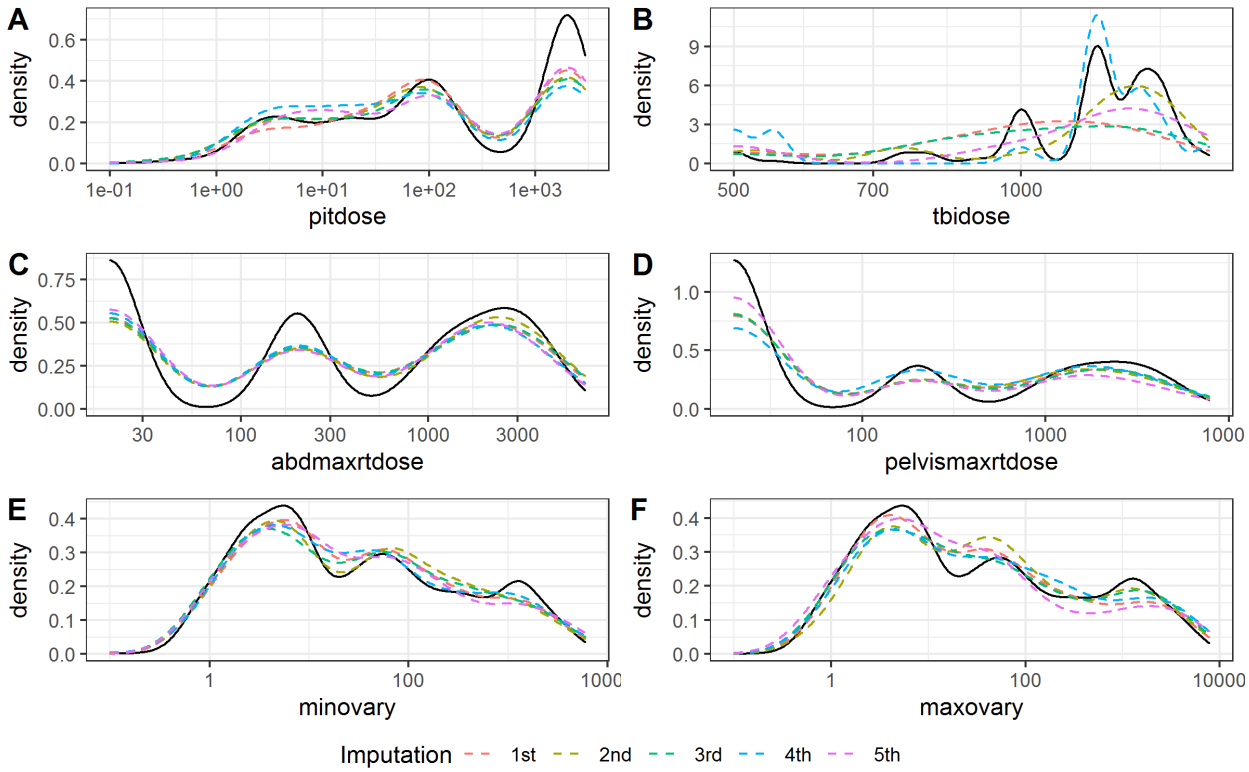


Figure 3.2 Kernel density estimates for the distributions of the radiation doses

Solid black line: observed data; dashed color lines: imputed data

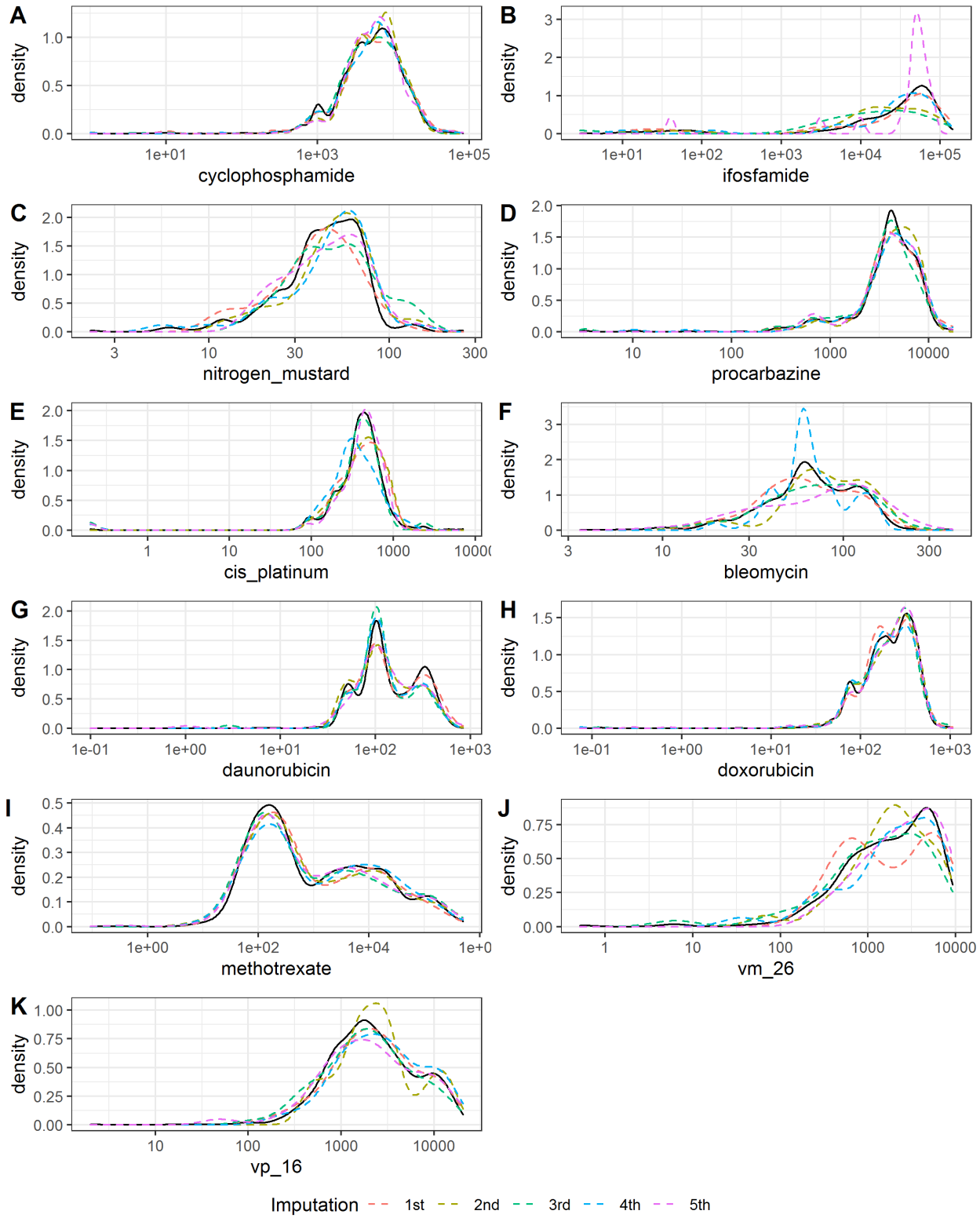


Figure 3.3 Kernel density estimates for the distributions of the chemotherapy agents' doses
Solid black line: observed data; dashed color lines: imputed data

Table 3.1 Summary of frequencies and proportions of missing value and non-zero value in continuous variables

Variable	Frequency of missing	Missing proportion	Frequency of non-zero values in observed data	Proportion of positive value in observed data	Frequency of positive values in imputed data	Proportion of non-zero value in imputed data
Age at event	38	0.5%	7853	100.0%	38	100.0%
Pituitary RT dose	675	8.6%	3469	48.1%	319	49.9%
Ovary RT dose (min)	709	9.0%	3448	48.0%	363	53.9%
Ovary RT dose (max)	715	9.1%	3442	48.0%	333	47.0%
TBI dose	644	8.2%	163	2.2%	22	3.7%
Abdomen RT dose	649	8.2%	3509	48.5%	306	46.2%
Pelvis RT dose	648	8.2%	3510	48.5%	306	48.1%
BCNU	491	6.2%	137	1.9%	9	3.3%
Busulfan	491	6.2%	61	0.8%	7	1.2%
CCNU	492	6.2%	63	0.9%	4	1.0%
Chlorambucil	484	6.1%	16	0.2%	1	0.2%
Cyclophosphamide	820	10.4%	2899	41.0%	369	46.6%
Ifosfamide	502	6.4%	338	4.6%	22	3.4%
Melphalan	488	6.2%	82	1.1%	7	2.5%
Nitrogen Mustard	530	6.7%	324	4.4%	32	6.6%
Procarbazine	603	7.6%	605	8.3%	111	16.4%
Thiotepa	487	6.2%	28	0.4%	8	2.1%
Carboplatin	501	6.3%	131	1.8%	12	2.4%
Cis-platinum	522	6.6%	452	6.1%	42	6.9%
Bleomycin	526	6.7%	561	7.6%	42	9.9%
Daunorubicin	543	6.9%	1128	15.4%	86	17.9%
Doxorubicin	700	8.9%	2654	36.9%	288	42.6%
Idarubicin	479	6.1%	58	0.8%	2	0.6%
Methotrexate	761	9.6%	2796	39.2%	283	35.6%
Mitoxantrone	485	6.1%	29	0.4%	3	0.4%
VM 26	500	6.3%	292	4.0%	27	4.0%
VP 16	568	7.2%	968	13.2%	82	15.5%

3.2 Inverse Probability Censoring Weights

3.2.1 Introduction

Censored observation is a key feature of the time to event data. The modeling of time to event data is often through accelerated failure time (AFT) models (such as Weibull, log-normal, log-logistic, etc), and semi-parametric models (Cox PH). The Cox PH model is the most popular method because it can incorporate multiple variables and does not depend on the distributional assumptions of the time variable required by the AFT models. However, Cox PH is typically used to establish the relationship between risk factors and time to event rather than to predict risk. In addition, it requires the proportional hazards assumption which may not hold in practice.[60]

An alternative way to deal with the censoring problem is employing the idea of inverse probability of censoring weighting (IPCW) proposed by Robins et al in the 1990s.[61] Using this method, we can take censoring into account by weighting observed outcomes. Then the age to event problem can be reframed to a problem with the binary outcome without losing the information of censored data, enabling the use of various classification algorithms, such as logistic regression and many popular machine learning algorithms.[62] The principle of IPCW has been well established[60]. The process of deriving IPCW for this study was described in the following section.

3.2.2 Derivation

The goal of this study is to estimate the probability of POI for a female childhood cancer survivor before a cut-off age a_0. For a female childhood cancer survivor with covariates z_i (a realization of potential predictors Z), let random variable A_i be the POI age and random variable C_i be the censoring age.

Define the observed age at event $X_i = min\{A_i, C_i\}$, censoring indicator $\delta_i = I(A_i < C_i)$, and censoring status $s_{a_0,i}$ at cut-off age a_0:

$$s_{a_0,i} = \begin{cases} 1, & if\ (X_i < a_0\ \&\ \delta_i = 1)\ or\ (X_i \geq a_0) \\ 0, & if\ (X_i < a_0\ \&\ \delta_i = 0) \end{cases}.$$

$s_{a_0,i} = 1$ represents the ovarian status is known at age a_0, $s_{a_0,i} = 0$ represents the ovarian status is censored at age a_0.

The probability of POI before a_0 can be expressed as:

$$P(A < a_0|Z) = f_{a_0}(Z, \theta),$$

where $f_{a_0}(Z, \theta)$ is the model for estimating the probability of POI before age a_0. θ are from parameters' space Θ. For parametric models (e.g., generalized linear regression models) θ are parameters, for non-parametric models (e.g. tree-based methods), θ indicate how the Z space is segmented.

Solving this problem equals to finding optimal θ, and it can be obtained by solving the following minimization problem:

$$\min_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^{N} \ell\left(f_{a_0}(z_i, \theta), y_i\right)$$

wherein N is the number of observations, y_i is the event outcome equals to I($A_i < a_0$) and $\ell(\cdot)$ is an optimization function, for example for logistic regressions

$$\ell\left(f_{a_0}(z_i, \theta), y_i\right) = y_i log\left(f_{a_0}(z_i, \theta)\right) + (1 - y_i) log\left(1 - f_{a_0}(z_i, \theta)\right).$$

Due to censoring, we cannot observe all the event outcomes, and if we ignore the unobserved data, the minimization problem will become:

$$\min_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^{N} s_{a_0,i}\, \ell\ \left(f_{a_0}(z_i, \theta), y_i\right).$$

This optimization problem leads to a misspecified θ and a model that can not be generalized to new data. To solve this problem, we can modify the above equation using the idea of inverse probability weighting, i.e., adding the probability of remaining uncensored at a_0 for each uncensored observation:

$$\min_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^{N} \frac{s_{a_0,i}}{P(s_{a_0,i})} \ell \left(f_{a_0}(z_i, \theta), y_i \right),$$

According to the definition of $s_{a_0,i}$, $P(s_{a_0,i})$ can be expressed as:

$$P(s_{a_0,i}) = \begin{cases} P(C_i > X_i | z_i, A_i), & if\ X_i < a_0,\ \delta_i = 1 \\ P(C_i > a_0 | z_i), & if\ X_i \geq a_0 \end{cases}$$

This modified optimization problem can give us a consistent solution for θ.

Proof:

$$E\left\{ \frac{s_{a_0,i}}{P(s_{a_0,i})} | z_i, A_i \right\} = E\left\{ \left[\frac{I(X_i \geq a_0)}{P(C_i > a_0 | z_i)} + \frac{I(X_i < a_0)\delta_i}{P(C_i > X_i | z_i, A_i)} \right] | z_i, A_i \right\} = 1$$

Here, we use the information from the entire sample to estimate $P(C_i > a_0 | z_i)$ and $P(C_i > X_i | z_i, A_i)$. To estimate $P(C_i > X_i | z_i, A_i)$ where A_i is not fully observed, we rely on an important assumption that the **censoring process $\mathbf{C_i}$ and event process $\mathbf{A_i}$ are independent conditional on covariates $\mathbf{z_i}$**. Under this assumption $P(C_i > X_i | \mathrm{z}_i, \mathrm{A}_i)$ is equal to $P(C_i > X_i | \mathrm{z}_i)$, enabling us to use observed covariates to build an estimator $\widehat{\mathrm{G}}(\cdot)$ for $P(s_{a_0,i})$. Then the estimated IPCW can be written as:

$$\widehat{\omega}_{a_0,i} = \frac{s_{a_0,i}}{P(s_{a_0,i})} = \frac{I\,(X_i < a_0)\delta_i}{\hat{G}(X_i | z_i)} + \frac{I\,(X_i \geq a_0)}{\hat{G}(a_0 | z_i)}$$

If the censoring process does not depend on any covariates, then $\widehat{\mathrm{G}}(\cdot)$ can be estimated by the Kaplan-Meier method. On the other hand, if the censoring process depends on any covariates z_i,

then the $\widehat{\mathrm{G}}(\cdot)$ can be estimated by a covariate-specific survival model including Cox PH models, AFT models, and random survival forest models.

3.2.3 Competing risk

Another challenge in this study is the presence of competing risks, i.e. SPM. Once a female experienced SPM, she was not at risk of POI. Therefore, we cannot treat them as censoring. To solve this problem, we modified the IPCW formula by adding a POI indicator p_i in the first term and counting SPM as uncensoring when estimating $\widehat{\mathrm{G}}(\cdot)$:

$$\widehat{\boldsymbol{\omega}}_{\boldsymbol{a_0},\boldsymbol{i}} = \frac{\boldsymbol{I(X_i < a_0)I(p_i = 1)\delta_i}}{\widehat{\boldsymbol{G}}(\boldsymbol{X_i})} + \frac{\boldsymbol{I(X_i \geq a_0)}}{\widehat{\boldsymbol{G}}(\boldsymbol{a_0})}$$

Therefore, in the denominator, SPM was not treated as censoring, so they did not increase the weights to the remaining survivors. On the other hand, due to the existence of POI indicator p_i in the first term, the weights for survivors with SPM ovarian status were assigned as zero, i.e., SPM did not count as an event like POI. In summary, the IPCW for different status and age at events/censoring were summarized in Table 3.2.

Table 3.2 Assignment of IPCW

$\mathrm{X_i}$	ovarian status	δ_i censoring indicator	p_i POI indicator	$\widehat{\omega}_{a_0,i}$
$\leq a_0$	Normal	0	0	0
	POI	1	1	$1/\widehat{\mathrm{G}}(\mathrm{X_i})$
	SPM	1	0	0
$> a_0$	Normal	0	0	$1/\widehat{\mathrm{G}}(a_0)$
	POI	1	0	$1/\widehat{\mathrm{G}}(a_0)$
	SPM	1	0	$1/\widehat{\mathrm{G}}(a_0)$

3.2.4 Results

Random survival forest models were fitted for estimating $\widehat{G}(\cdot)$. The variables and their importance in the survival model were presented in Figure 3.4. Age at diagnosis ("age_dx") was the most important predictor for estimating $\widehat{G}(\cdot)$.

Figure 3.5 showed the distribution of estimated IPCW when the age threshold increased from 21 to 39. As the proportion of censoring increased with the age threshold, individuals were assigned higher IPCW and more patients were weighted to be 0 (i.e., censored). The highest IPCW was close to 30 at the largest age threshold of 39.

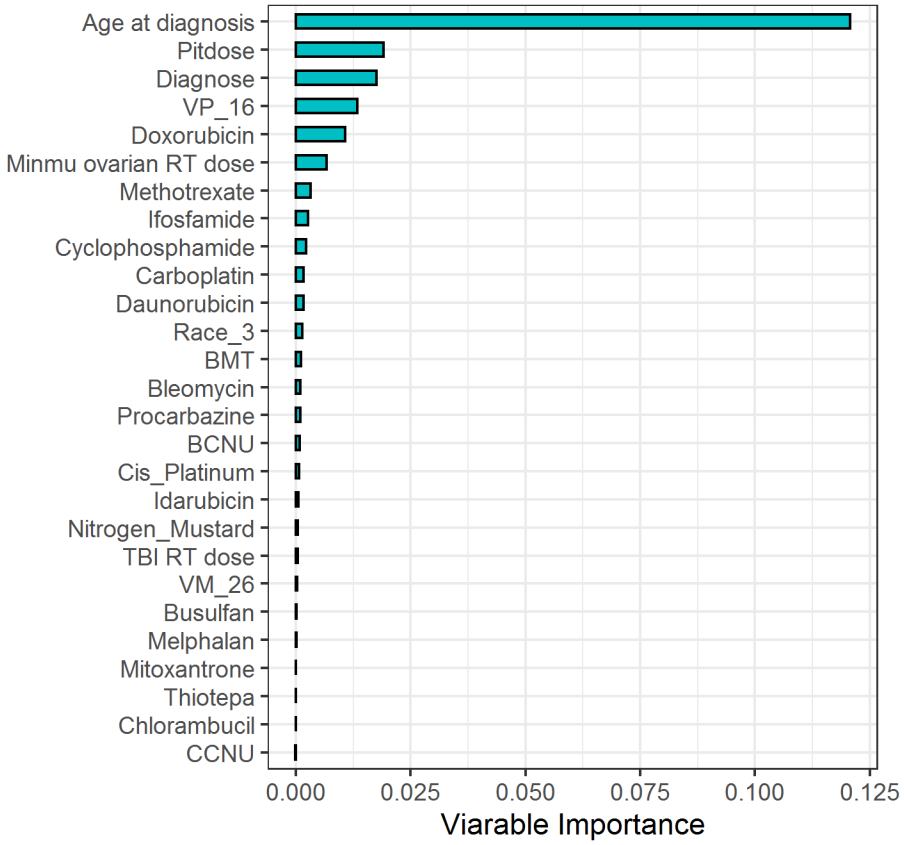


Figure 3.4 Variable importance in the random survival forest model for censoring

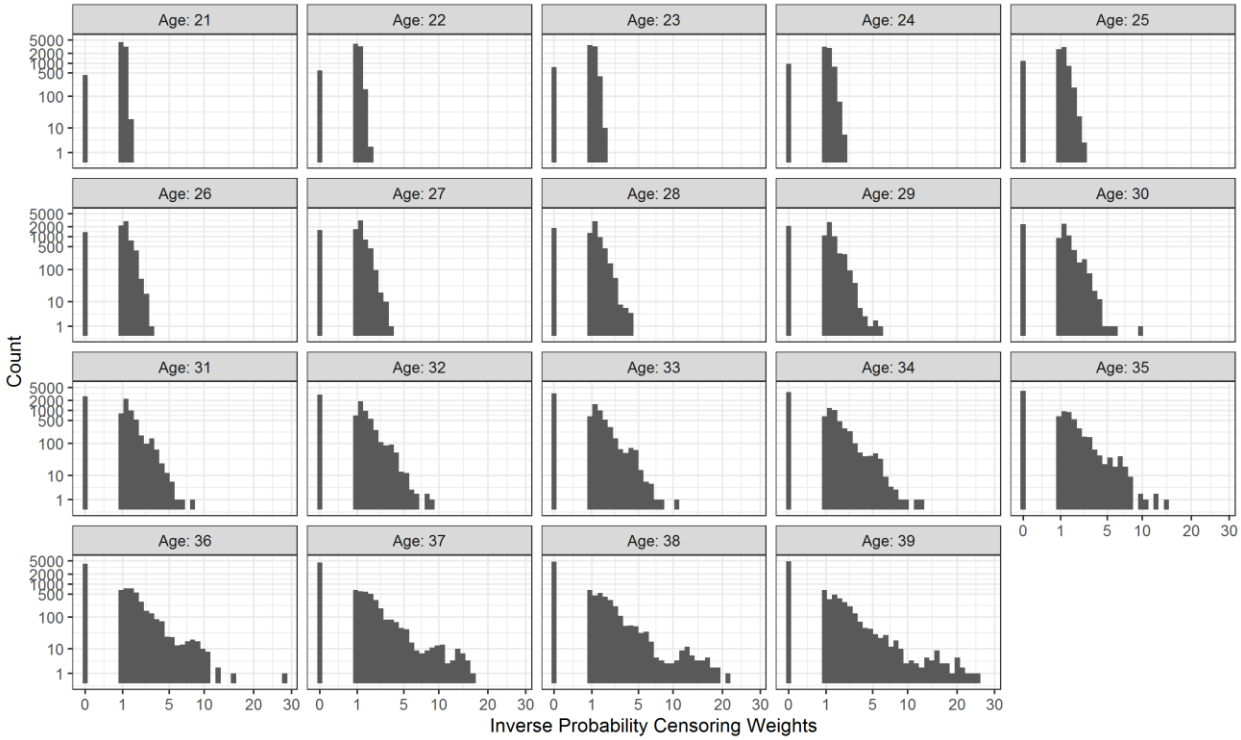


Figure 3.5 Distribution of the weights at different age thresholds ranging from 21 to 39

4 Model development and evaluation

This chapter focused on model development and evaluation. Section 4.1 specified the outcomes and candidate predictors for model development. Section 4.2 introduced the three algorithms for model development and section 4.3 highlighted the procedure of evaluating model performance. Sections 4.4 and 4.5 presented the results and discussions.

.

4.1 Outcomes and predictors

To predict the risk of developing POI by age threshold ranging from 21 to 39, the outcomes were defined as whether a female CCS developed POI by age 21, 22, …, and 39, accordingly. The proportions of components of outcomes by the nineteen different age thresholds were listed in Table 4.1. As the age threshold increased from 21 to 39, the proportion of subjects with censored ovarian status increased significantly from 5.3% to 60.1%.

The candidate predictors included race, cancer diagnosis type, age at diagnosis, radiation dose, chemotherapy agents, and BMT. Race and cancer diagnosis type were categorical variables. For the ten alkylating agents and their derived cumulative dose CED, the information would be redundant if included all of them in one model. Therefore, two sets of variables were prepared separately; each of them considered either ten individual alkylating agents or CED. A summary of the candidate predictors was listed in Table 4.2.

Table 4.1 Ovarian status distribution at different age thresholds

Cut-off age years	Outcomes n (%)			Censored n (%)	Sample size n
	Normal	POI	SPM		
21	6809 (86.4)	607 (7.7)	13 (0.2)	416 (5.3)	7883
22	6605 (83.8)	639 (8.1)	16 (0.2)	582 (7.4)	7880
23	6417 (81.5)	658 (8.4)	18 (0.2)	747 (9.5)	7878
24	6208 (78.8)	680 (8.6)	19 (0.2)	931 (11.8)	7876
25	5968 (75.8)	703 (8.9)	23 (0.3)	1140 (14.5)	7872
26	5707 (72.6)	718 (9.1)	28 (0.4)	1374 (17.5)	7865
27	5475 (69.7)	727 (9.3)	40 (0.5)	1575 (20.1)	7855
28	5201 (66.3)	736 (9.4)	47 (0.6)	1824 (23.2)	7846
29	4925 (62.8)	748 (9.5)	60 (0.8)	2072 (26.4)	7843
30	4617 (58.9)	756 (9.7)	72 (0.9)	2350 (30)	7833
31	4296 (54.9)	782 (10)	91 (1.2)	2619 (33.5)	7826
32	3987 (51)	793 (10.1)	111 (1.4)	2891 (37)	7820
33	3667 (46.9)	813 (10.4)	135 (1.7)	3162 (40.5)	7815
34	3372 (43.2)	829 (10.6)	148 (1.9)	3422 (43.8)	7809
35	3034 (38.9)	841 (10.8)	173 (2.2)	3716 (47.6)	7802
36	2715 (34.9)	864 (11.1)	197 (2.5)	3975 (51)	7789
37	2424 (31.1)	878 (11.3)	220 (2.8)	4223 (54.3)	7783
38	2127 (27.4)	888 (11.4)	234 (3)	4488 (57.7)	7775
39	1894 (24.4)	905 (11.7)	258 (3.3)	4668 (60.1)	7763

Table 4.2 Candidate predictors considered during model development

Type of variables	Candidate predictors setting 1	Candidate predictors setting 2	
Categorical	*Race* (3 levels) white(reference), black, and other	*Race* (3 levels) white(reference), black, and other	
	diagnosis (8 levels) leukemia(reference), central nervous system cancers, neuroblastoma, non-Hodgkin lymphoma, Hodgkin lymphoma, kidney cancer, bone tumors, soft-tissue sarcoma	*diagnosis* (8 levels) leukemia(reference), central nervous system cancers, neuroblastoma, non-Hodgkin lymphoma, Hodgkin lymphoma, kidney cancer, bone tumors, soft-tissue sarcoma	
	BMT (Yes/No)	*BMT* (Yes/No)	
Continuous: Irradiation dose (Gy)	*total body irradiation dose*	*total body irradiation dose*	
	minimum ovary radiation dose	*minimum ovary radiation dose*	
	radiation dose to pituitary	*radiation dose to pituitary*	
Continuous: Alkylating agents' doses (g/m^2)	*CED*	*BCNU*	*Busulfan*
		CCNU	*Chlorambucil*
		Cyclophosphamide	*Ifosfamide*
		Melphalan	*Nitrogen Mustard*
		Procarbazine	*Thiotepa*
Continuous: Other chemotherapy agents' doses (g/m^2)	*Carboplatin*	*Carboplatin*	
	Cis_Platinum	*Cis_Platinum*	
	Bleomycin	*Bleomycin*	
	Daunorubicin	*Daunorubicin*	
	Doxorubicin	*Doxorubicin*	
	Epirubicin	*Epirubicin*	
	Idarubicin	*Idarubicin*	
	Methotrexate	*Methotrexate*	
	Mitoxantrone	*Mitoxantrone*	
	VM 26	*VM 26*	
	VP 16	*VP 16*	

4.2 Algorithms for model development

Two machine learning algorithms: Elastic-Net panelized age-specific logistic regression[63] (EN-ALR) and XGBoost[64] were used to mapping predictors to outcomes (details about the algorithms were available in Appendix G). The third algorithm "Ensemble" averaged the predicted risks from the previous two.

The candidate predictors (Table G.1 in Appendix G) in EN-ALR and XGBoost were the same except for some rarely used chemotherapy agents (i.e. busulfan, CCNU, chlorambucil, melphalan, thiotepa, idarubicin, and mitoxantrone). These seven chemotherapy agents were coded as binary Yes/No in EN-ALR but were retained as continuous variables in XGBoost as the algorithm automatically choose the split-point to maximize the information gain.

Both EN-ALR and XGBoost select predictors automatically through tuning hyperparameters. As a result, once the hyperparameters are selected, the model is fixed accordingly. Therefore, tuning hyperparameters is a key step that decides the performance of the final model. In this study, I employed the "random search"[65] approach to tune hyperparameters in a prespecified hyperparameter space (Appendix G).

4.3 Procedures for model evaluation

Typically, model evaluation and hyperparameter tunning are conducted in the same CV process. However, this evaluation procedure may be overoptimistic about the model performance because the hyperparameters tuned in the training set were chosen based on the model performance in validation sets. To address this issue, I employed a nested CV to evaluate the performance of the modeling procedure. The rationality of nested CV was that it constructed two layers of CV: the

inner CV was used for tunning hyperparameters and the outer CV was served for evaluating model performance (Appendix E).

In summary, I used a classical CV to obtain the optimal hyperparameters and a nested CV to evaluate the performance of the modeling procedure (rather than the final model). The whole modeling process was illustrated in a schematic diagram (Figure 4.1).

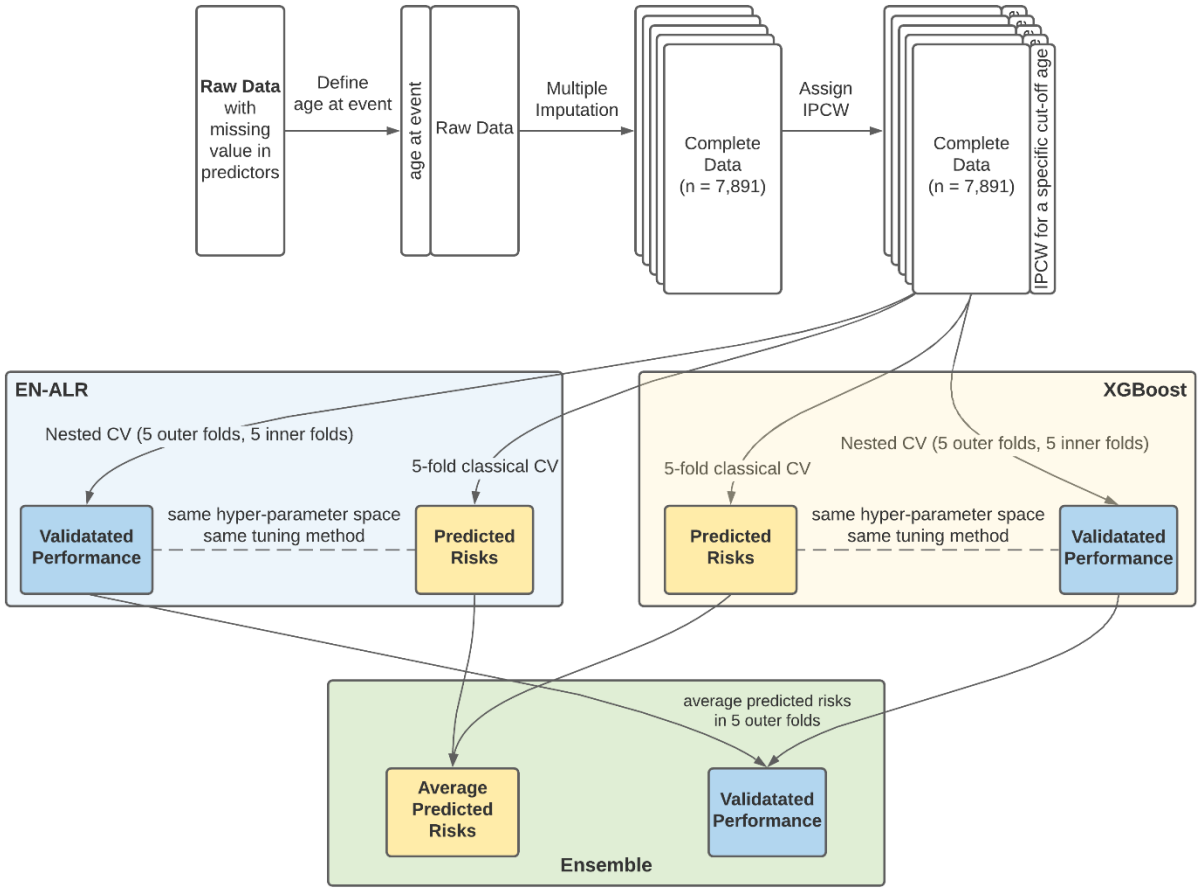


Figure 4.1 Schematic diagram of the modeling procedure to generate predicted risks

It should be noted that combining MI and nested CV generated 125 different data sets (see Chapter 3.1.3); therefore, for each hyperparameter setting, there were 125 models developed to predict the risk of POI, and the average of the 125 predicted risks on one subject was used as her final predicted risk.

The performance matrices used in this research include areas under the receiver operating characteristic curves (AUC) and average positive predictive value (AP) for measuring the ability of discriminatory and prospective prediction. Scaled Brier Score (sBrS) was used to describe the overall model performance.

In addition, calibration curves that compared cumulative weighted predicted risk and cumulative weighted events were used to visually inspect the calibration of models.

4.4 Results

The results from nested CV showed that the performance of models with ten individual alkylating agents was superior to the models with CED (Appendix H). Therefore, the models with ten individual alkylating agents will be used and presented in the remaining sections in this chapter.

4.4.1 Predictors

Figure 4.2 illustrated the predictors selected by EN-ALR and XGBoost when the age threshold was 24 and 29, respectively. The boxplots in the left panel showed the coefficients of each variable in the 125 EN-ALR models and the boxplots in the right panel showed the variable importance in the 125 XGBoost models. If a variable was not selected by any of the 125 models, the corresponding boxplot was absent. For example, rarely used chemotherapy agents such as busulfan, CCNU, chlorambucil, melphalan, thiotepa, idarubicin, and mitoxantrone were not selected by XGBoost. Therefore, no boxplots were available for these predictors in the right panel. The width of a boxplot reflected the variation in coefficients or variable importance of the corresponding predictor across the 125 data sets.

The frequencies of the predictors selected by the 125 models as well as their range of values were shown in Table 4.3 and Table 4.4 for ages 24 and 29, respectively. For better visualization of the coefficients of radiotherapy and chemotherapy, the unit of irradiation doses was set to be Gy, and the unit of chemotherapy agent doses was set to be g/m^2.

Based on results from the XGBoost algorithm (right panels in Figure 4.2), minimum ovarian radiation dose and BMT were the top two risk predictors in estimating the risk of POI by both ages 24 and 29. These two predictors had positive coefficients in EN-ALN models, indicating patients treated with BMT and who received higher minimum ovarian radiation doses were at a greater risk of developing POI.

The contributions of twenty chemotherapy agents including the ten alkylating agents were examined individually. Cyclophosphamide, procarbazine, methotrexate, and VP 16 were chosen by both algorithms, indicating they were important variables for predicting POI. Bleomycin was also identified in algorithms for predicting POI by age 24. However, it had a negative adjusted coefficient, indicating higher doses of bleomycin were associated with reduced POI risk after adjusting for other variables.

In terms of race, “Black” and “Other” had positive coefficients comparing to “White”, indicating the two groups had a higher risk of developing POI than “White”.

As for the cancer types, the results in EN-ALR showed that while patients with CNS cancer and kidney tumors had higher risks than patients with leukemia, patients diagnosed with non-Hodgkin lymphoma and bone cancer had lower risks than patients with leukemia.

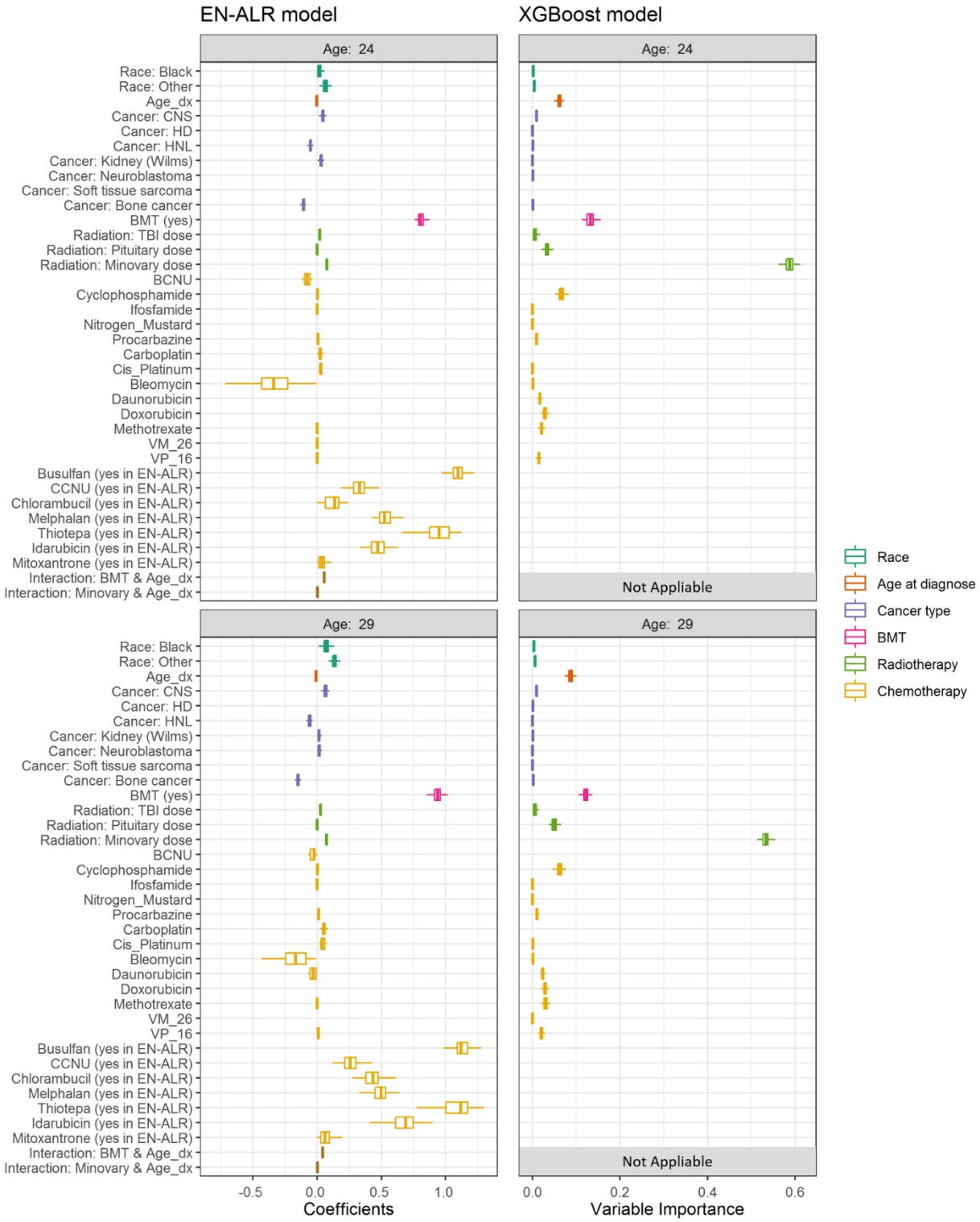


Figure 4.2 Predictors in the two algorithms
Left: coefficients in EN-ALR; Right: variable importance in XGBoost

Table 4.3 Coefficients in EN-ALR and Variable importance in XGBoost at Age 24. Proportion indicates the rates of each variable selected by algorithms in the 125 data sets. RT: radiotherapy; INT: interaction

Age 24		EN-ALR coefficients		XGBoost variable importance	
Variables		Proportion (%)	Median [min, max]	Proportion (%)	Median [min, max]
Race	Black	61.6	0.016 [0.001,0.077]	64.0	0.002 [<0.001,0.004]
	Other	100.0	0.063 [0.018,0.123]	100.0	0.004 [0.003,0.007]
Age	Age at diagnosis	100.0	-0.004 [-0.004,-0.002]	100.0	0.061 [0.047,0.079]
Cancer diagnosis type	CNS	99.2	0.046 [0.005,0.072]	100.0	0.009 [0.007,0.014]
	HD	0	NA	42.6	<0.001 [<0.001,0.003]
	HNL	100.0	-0.052 [-0.08,-0.028]	8.8	0.001 [0.001,0.001]
	Kidney (Wilms)	96.8	0.033 [0.004,0.068]	56.0	<0.001 [<0.001,0.003]
	Neuroblastoma	0	NA	84.8	0.001 [<0.001,0.003]
	Soft tissue sarcoma	0	NA	0	NA
	Bone cancer	100.0	-0.107 [-0.141,-0.087]	72.8	0.001 [<0.001,0.002]
BMT	BMT (yes)	100.0	0.806 [0.729,0.874]	100.0	0.133 [0.113,0.17]
	TBI dose	100.0	0.021 [0.008,0.033]	86.4	0.004 [0,0.031]
RT (Gy)	Pituitary dose	0.8	<0.001 [<0.001, <0.001]	100.0	0.033 [0.02,0.052]
	(min) ovarian dose	100.0	0.075 [0.069,0.079]	100.0	0.588 [0.561,0.618]
Chemotherapy doses (g/m^2)	BCNU	1.6	-0.078 [-0.118,-0.039]	0	NA
	Cyclophosphamide	99.2	0.003 [<0.001,0.008]	100.0	0.066 [0.052,0.087]
	Ifosfamide	2.4	<0.001 [<0.001, <0.001]	34.4	<0.001 [<0.001,0.001]
	Nitrogen Mustard	0	NA	0.8	<0.001 [<0.001, <0.001]
	Procarbazine	89.6	0.006 [<0.001,0.02]	100.0	0.009 [0.003,0.015]
	Carboplatin	94.4	0.024 [<0.001,0.064]	0	NA
	Cis_Platinum	1.6	0.027 [0.013,0.041]	11.2	<0.001 [<0.001,0.002]
	Bleomycin	94.4	-0.338 [-0.805,-0.002]	92.8	0.001 [0,0.005]
	Daunorubicin	0	NA	100.0	0.017 [0.011,0.023]
	Doxorubicin	0	NA	100.0	0.029 [0.02,0.04]
	Methotrexate	99.2	<0.001 [-0.001,<0.001]	100.0	0.022 [0.013,0.035]
	VM_26	5.6	-0.001 [-0.006,<0.001]	0	NA
	VP_16	5.6	0.001 [<0.001,0.006]	100.0	0.015 [0.007,0.024]
	Busulfan	100.0	1.097 [0.945,1.225]	0	NA
	CCNU	100.0	0.331 [0.117,0.484]	0	NA
	Chlorambucil	23.2	0.136 [0,0.242]	0	NA
	Melphalan	100.0	0.525 [0.422,0.707]	0	NA
	Thiotepa	100.0	0.949 [0.66,1.125]	0	NA
	Idarubicin	100.0	0.472 [0.238,0.696]	0	NA
	Mitoxantrone	26.4	0.034 [0.001,0.183]	0	NA
INT	BMT & Age at diagnosis	100.0	0.055 [0.048,0.06]	NA	NA
	Minovary & Age at diagnosis	100.0	0.002 [0.002,0.003]	NA	NA

Table 4.4 Coefficients in EN-ALR and Variable importance in XGBoost at Age 29. Proportion indicates the percentages of each variable was selected in the 125 data sets. RT: radiotherapy; INT: interaction

Age 24		EN-ALR coefficients		XGBoost variable importance	
Variables		Proportion (%)	Median [min, max]	Proportion (%)	Median [min, max]
Race	Black	100.00	0.072 [0.012,0.144]	100.0	0.003 [<0.001,0.006]
	Other	100.00	0.135 [0.082,0.188]	100.0	0.006 [0.004,0.009]
Age	Age at diagnosis	100.00	-0.009 [-0.01,-0.007]	100.0	0.086 [0.069,0.104]
Cancer diagnosis type	CNS	100.00	0.065 [0.012,0.098]	100.0	0.009 [0.006,0.012]
	HD	0	NA	95.2	0.001 [<0.001,0.004]
	HNL	100.00	-0.06 [-0.1,-0.033]	100.0	<0.001 [<0.001,0.002]
	Kidney (Wilms)	66.40	0.015 [0.001,0.049]	99.2	0.001 [<0.001,0.004]
	Neuroblastoma	84.80	0.016 [<0.001,0.056]	95.2	<0.001 [<0.001,0.002]
	Soft tissue sarcoma	0	NA	2.4	<0.001 [<0.001,<0.001]
	Bone cancer	100.00	-0.15 [-0.188,-0.125]	100.0	0.002 [<0.001,0.005]
BMT	BMT (yes)	100.00	0.939 [0.853,1.054]	100.0	0.121 [0.106,0.15]
	TBI dose	100.00	0.028 [0.015,0.04]	91.2	0.005 [<0.001,0.023]
RT (Gy)	Pituitary dose	18.40	<0.001 [<0.001,0.001]	100.0	0.05 [0.037,0.065]
	Minovary dose	100.00	0.073 [0.067,0.077]	100.0	0.534 [0.513,0.555]
Chemotherapy doses (g/m^2)	BCNU	5.60	-0.021 [-0.154,<0.001]	0	NA
	Cyclophosphamide	100.00	0.004 [0.001,0.009]	100.0	0.062 [0.045,0.077]
	Ifosfamide	12.80	<0.001 [<0.001,0.001]	68.0	<0.001 [<0.001,0.001]
	Nitrogen Mustard	0	NA	40.8	0.001 [<0.001,0.002]
	Procarbazine	97.60	0.011 [0.001,0.023]	100.0	0.011 [0.006,0.017]
	Carboplatin	99.20	0.052 [0.002,0.104]	0	NA
	Cis_Platinum	2.40	0.039 [0.017,0.074]	12.8	0.001 [<0.001,0.003]
	Bleomycin	41.60	-0.166 [-0.431,-0.009]	93.6	0.001 [<0.001,0.004]
	Daunorubicin	10.40	-0.031 [-0.071,<0.001]	100.0	0.023 [0.017,0.031]
	Doxorubicin	0	NA	100.0	0.029 [0.02,0.038]
	Methotrexate	100.00	-0.001 [-0.001,<0.001]	100.0	0.03 [0.02,0.041]
	VM_26	0	NA	55.2	0.001 [<0.001,0.002]
	VP_16	96.00	0.008 [<0.001,0.023]	100.0	0.021 [0.013,0.033]
	Busulfan	100.00	1.121 [0.904,1.275]	0	NA
	CCNU	100.00	0.257 [0.001,0.531]	0	NA
	Chlorambucil	98.40	0.433 [0.029,0.669]	0	NA
	Melphalan	100.00	0.498 [0.33,0.658]	0	NA
	Thiotepa	100.00	1.116 [0.645,1.3]	0	NA
	Idarubicin	100.00	0.69 [0.386,0.957]	0	NA
	Mitoxantrone	66.40	0.061 [0.001,0.273]	0	NA
INT	BMT & Age at diagnosis	100.00	0.043 [0.037,0.05]	NA	NA
	Minovary & Age at diagnosis	100.00	0.002 [0.002,0.003]	NA	NA

4.4.2 Model Performance

The nested CV evaluated AUC, AP, and sBrS for the three algorithms were shown in Table 4.5. The point estimates of AUC ranged from 0.776 to 0.795 in the models for age 24 and from 0.771 to 0.791 in the models for age 29. The AP ranged from 0.464 to 0.480 in the models for age 24 and from 0.473 to 0.495 in the models for age 29. The sBrS ranged from 0.238 to 0.264 in the models for age 24 and from 0.230 to 0. 0.259 in the models for age 29.

Table 4.5 Nested CV evaluated performance at age 24 and 29

Performance		**Age: 24**		**Age: 29**	
		Point Estimate	**95% CI[a]**	**Point Estimate**	**95% CI[a]**
AUC	EN-ALR	0.776	(0.754, 0.798)	0.771	(0.750, 0.791)
	XGBoost	0.780	(0.770, 0.811)	0.787	(0.767, 0.807)
	Ensemble	0.795	(0.775, 0.817)	0.791	(0.771, 0.811)
AP	EN-ALR	0.464	(0.426, 0.507)	0.473	(0.433, 0.515)
	XGBoost	0.470	(0.428, 0.512)	0.482	(0.437, 0.522)
	Ensemble	0.480	(0.440, 0.522)	0.495	(0.454, 0.534)
sBrS	EN-ALR	0.238	(0.209, 0.266)	0.230	(0.200, 0.258)
	XGBoost	0.262	(0.226, 0.296)	0.255	(0.216, 0.288)
	Ensemble	0.264	(0.233, 0.294)	0.259	(0.226, 0.289)
Event rate		0.089	(0.083, 0.096)	0.105	(0.099, 0.112)

a: 95% CI was calculated by the “Bootstrap” method.

XGBoost and Ensemble provided comparable values of AUC, AP, sBrS which were always higher than that of EN-ALR regardless of age. Between XGBoost and Ensemble, Ensemble presented slightly better performance. This pattern remained the same across different ages from 21 to 39 (shown in Figure 4.3). Overall, the Ensemble algorithm achieved the best performance among the

three algorithms at different ages: its AUCs were around 0.8 (ranged from 0.785 at age 31 to 0.801 at age 34), AP increased from 0.469 at age 21 to 0.595 at age 39 as event rates increased (from 0.079 at age 21 to 0.173 at age 39), and sBrS ranged from 0.259 at age 29 to 0.292 at age 37.

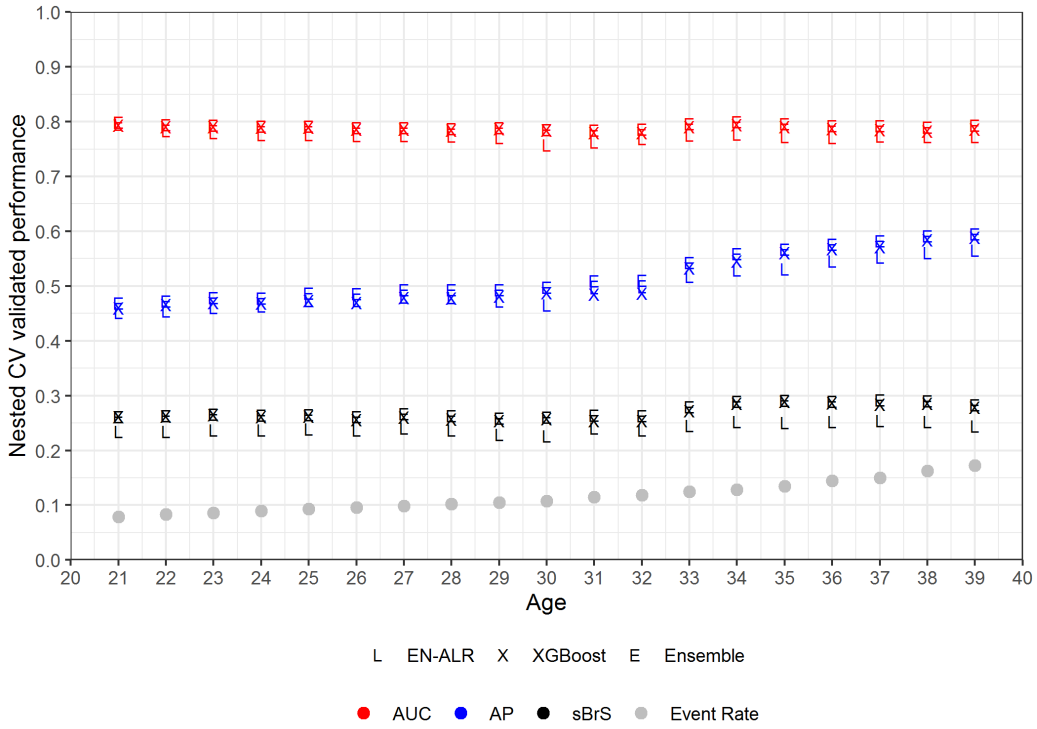


Figure 4.3 AUC, AP, sBrS at ages from 21 to 39

Figure 4.4 showed the calibration curves for the three algorithms at the age threshold from age 21 to 39. The calibration curves before age 28 followed the diagonal line well, indicating that the predicted risk had good alignment with the observed events. However, after age 28, the calibration curves started to deviate from the diagonal line. Serious deviations were presented after age 30, suggesting that the models were not well calibrated for ages over 30.

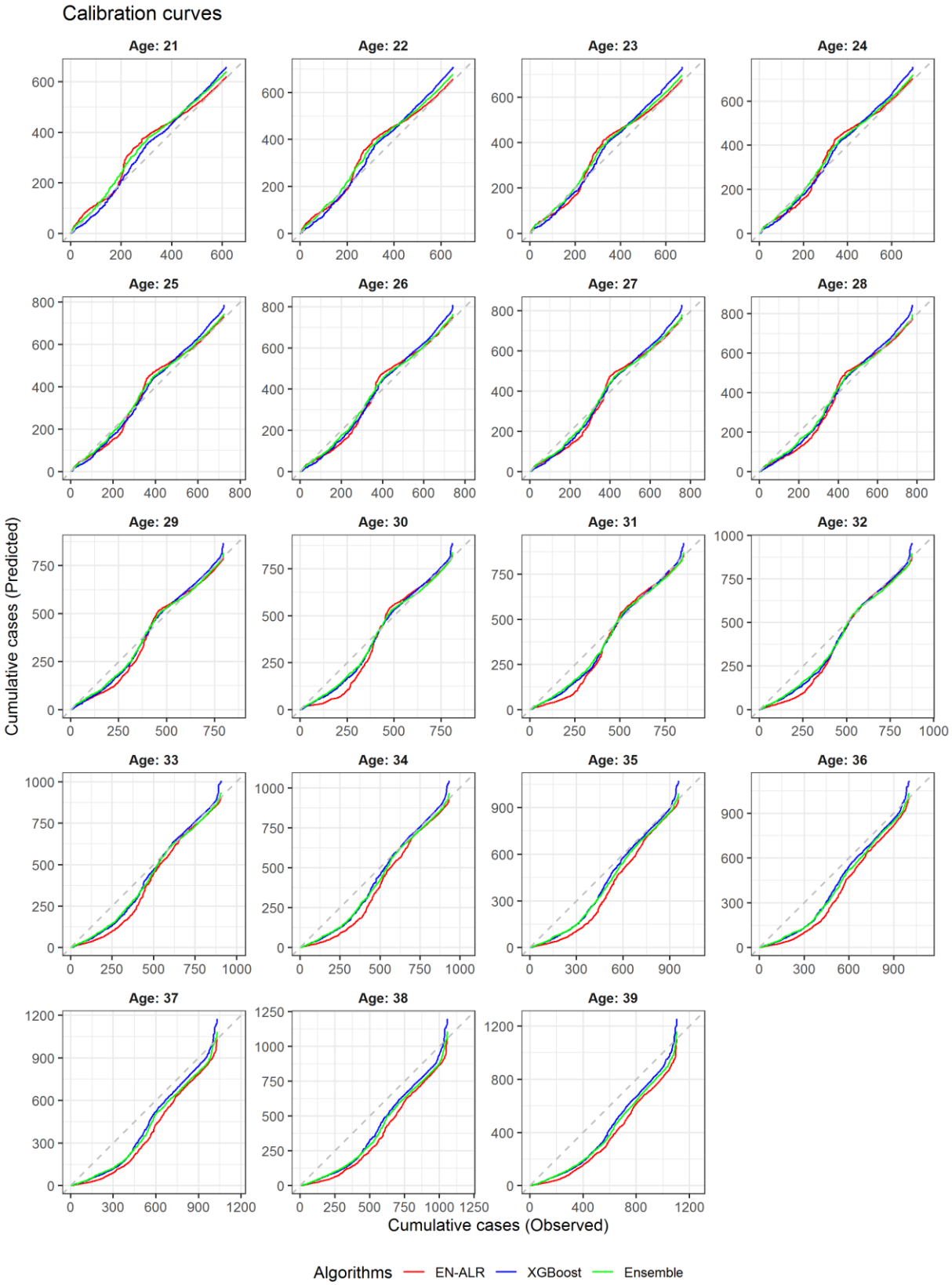


Figure 4.4 Calibration curves from EN-ALR, XGBoost, and Ensemble for different ages

4.4.3 Predicted risks

As the Ensemble algorithm achieved the best-validated performance, it was used to predict the risk of age-specific POI in the whole data set. It should be noted that the final predicted risks for individuals were obtained by averaging the predicted risks from the 125 work data sets.

Based on the suggestions from endocrinologists and pediatric oncologists, the predicted risks were stratified into four categories: <5%, 5% to <20%, 20% to <50%, and ⩾50%, representing low, medium-low, medium, and high-risk groups, respectively. Table 4.6 (all the numbers were weighted with IPCW weights) illustrated how the Ensemble algorithm categorized survivors into four categories.

Specifically, at age 24, 3495 (44.9%) of 7786 participants were estimated to be at low risk (52 [1.5%] developed POI), whereas 290 (3.7%) individuals were estimated to be at high risk (231 [79.7%] developed POI). At age 29, 1423 (18.9%) of 7533 participants were estimated to be at low risk (11 [0.8%] developed POI), whereas 341 (4.5%) individuals were estimated to be at high risk (280 [82.1%] developed POI). The results suggest that the Ensemble algorithm can successfully distinguish between survivors with low-risk and high risk.

Table 4.6. POI categories and prevalence for each cohort as predicted by the Ensemble algorithm (%: row percentage)

Predicted Risk	Age 24		Age 29	
	Survivors	POI (%)	Survivors	POI (%)
<5%	3495	51 (1.5%)	1423	11 (0.8%)
5% to <20%	3770	277 (7.4%)	5474	370 (6.8%)
20% to < 50%	330	150 (44.5%)	294	139 (47.3%)
≥50%	290	231 (79.6%)	341	280 (82.1%)
Total	7786	709 (9.1%)	7533	800 (10.6%)

4.5 Discussion

In line with the established risk factors in the literature, the results from both EN-ALR and XGBoost showed that BMT, minimum ovarian radiation dose, cyclophosphamide dose, and procarbazine dose were associated with the risk of developing POI. EN-ALR also identified the age at diagnosis as an effect modifier of BMT which was consistent with the findings of Clark (2020)[9]. Besides, both algorithms identified race *black* and *other* might have a higher risk of developing POI than *white*, which has not been well recognized in previous research. Therefore, although the predictors were automatically selected, they may give some insight into investigating the risk factors of developing POI. **However, it should be noted that these predictors were chosen because they can improve prediction accuracy, which, however, does not imply that they cause POI.** To conclude a causal relationship, a different research path is needed.

This research carefully designed the procedure of model evaluation to avoid the issue of overfitting and overoptimism. The nested CV results showed AUC could reach as high as 0.8, indicated that the models could well discriminate the subjects with POI from those without POI. The AP results were much higher than the population event rate, indicating a strong predictive power for detecting POI. The results of calibration curves showed a good alignment between predicted risks and observed events when the age threshold was less than 28, indicating that the models could well predict the probabilities of developing POI at a younger age.

The calibration results showed that long-term risk prediction can be challenging. One possible reason is that a large proportion of censoring is presented at an older age. Another reason might be that when the age threshold was farther away from the age at diagnosis, the effects of the

environment or the survivors' lifestyle may have come to play in the development of POI, which was not considered in this research.

As for the application of the final models, female survivors can be stratified into four risk categories according to estimated risks, providing useful information for them and clinicians to discuss their need for fertility preservation. Furthermore, The developed algorithm can be crafted into a user-friendly clinical tool. Appendix I presented two examples of using the tool to predict the risks of developing POI at different ages for patients.

5 Conclusions

This research aimed to develop accurate prognostic models for predicting absolute risks for an individual childhood cancer survivor developing POI by prespecified ages. The problem of missing data, censoring, and overfitting was carefully addressed. The validated model performance confirmed that the developed models can discriminate well between those survivors who developed POI and those survivors who did not, and the models can provide an accurate absolute estimated risk of developing POI before age 28.

In this chapter, I summarized the contributions of this thesis (section 5.1), discussed the limitations (section 5.2), and made recommendations on future work (section 5.3).

5.1 Summary

In Chapter 2, the data from CCSS was cleaned and explored. A new variable, *age at event,* was derived based on ovarian status and other menstrual history information. To ensure the accuracy of the outcome variable, I discussed the algorithm for deriving it extensively with a pediatric endocrinologist.

The problem of missing data was addressed in Chapter 3 by employing multiple imputation. The details of imputation including imputation model selection, iteration number determination, and post-processing were described. Furthermore, special consideration was given to how to implement multiple imputation and model performance evaluation properly together so that the information in validation sets did not 'leak' to the training sets.

Chapter 3 also addressed the problem of censoring by assigning individuals with inverse probability censoring weights. This method took into account censored subjects by assigning

weights of greater than 1 to those with observed ovarian status at a prespecified age. Since the censoring process was associated with covariates, a random survival forest was used to calculate the probability of remaining uncensored. The competing risk was considered in the formula of weights.

In Chapter 4, two modern machine learning algorithms: EN-ALR and XGBoost were employed and an "Ensemble" algorithm was used to take advantage of the two previous algorithms. The hyperparameter tuning strategy "random search" was used to find optimal hyperparameter settings. To avoid over-optimistic about the model performance, nested CV was employed to give an honest evaluation. An "Ensemble" method achieved the best performance. Its good discriminative power and calibration results (when the age threshold was less than 28) suggested that the final models could be used to predict POI in new data.

5.2 Limitations

Approximately 16% of the female CCSS participants (Figure 2.2) did not complete a questionnaire that contained the menstrual history section. In this research, we assumed those who failed to participate in the surveys had a similar pattern of POI to those who participated. However, a risk of bias would arise if the reason for not participating was associated with the menstrual status.

As a retrospective cohort study, CCSS sent out surveys containing menstrual history sections almost every seven years. This means many participants had to recall their health condition many years ago, implying a risk of recall bias. Especially, when individuals recalled the age of stopping

menstruating which happened many years ago, an inaccurate age might be reported. This would influence the outcomes we used to build the models.

The validity of using the multiple imputation and IPCW methods rely on the assumption of missing at random and the assumption of independence between the event process and censoring process given observed covariates. Although the two assumptions are reasonable in this research, it cannot be proved because the missing predictors and censored outcomes are unobserved.

In terms of the final model performance, although AUC, AP showed good performance, calibration curves started deviating from the diagonal line when age was greater than 28, implying that the estimated risk of developing POI by ages over 28 needs to be improved.

5.3 Future work

Future work could be beneficial when more data is released from CCSS. This might alleviate the problem of censoring and thus improving the long-term risk prediction.

The performance could be improved by tuning more hyperparameter settings. Especially for the XGBoost model, many more hyperparameters could be tuned. Furthermore, by using the "Ensemble" method, some other machine learning algorithms such as neural networks, support vector machines, and random forest could also contribute to improving ensemble performance.

Finally, although the model performance was carefully evaluated in this research, it may not reflect its performance in other childhood cancer populations. An external validation study would be ideal in the future.

Bibliography

1. Wilkosz P, Greggains GD, Tanbo TG, Fedorcsak P. Female Reproductive Decline Is Determined by Remaining Ovarian Reserve and Age. *PLoS One*. 2014;9(10). doi:10.1371/journal.pone.0108343

2. Chemaitilly W, Li Z, Krasin MJ, 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):2242-2250.

3. Levine JM, Whitton JA, Ginsberg JP, 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;124(5):1044-1052. doi:10.1002/cncr.31121

4. Chemaitilly W, Mertens AC, Mitby P, et al. Acute Ovarian Failure in the Childhood Cancer Survivor Study. *The Journal of Clinical Endocrinology & Metabolism*. 2006;91(5):1723-1728. doi:10.1210/jc.2006-0020

5. Torrealday S, Pal L. Premature Menopause. *Endocrinology and metabolism clinics of North America*. 2015;44(3):543-557.

6. Primary Ovarian Insufficiency | Hormone Health Network. https://www.hormone.org/diseases-and-conditions/primary-ovarian-insufficiency. Accessed November 17, 2020.

7. Oktay K, Harvey BE, Partridge AH, et al. Fertility Preservation in Patients With Cancer: ASCO Clinical Practice Guideline Update. *JCO*. 2018;36(19):1994-2001. doi:10.1200/JCO.2018.78.1914

8. Dudzinski DM. Ethical issues in fertility preservation for adolescent cancer survivors: oocyte and ovarian tissue cryopreservation. *Journal of Pediatric and Adolescent Gynecology*. 2004;17(2):97-102. doi:10.1016/j.jpag.2004.01.004

9. Clark RA, Mostoufi-Moab S, Yasui Y, 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;21(3):436-445.

10. Robison LL, Mertens AC, Boice JD, et al. Study design and cohort characteristics of the childhood cancer survivor study: A multi-institutional collaborative project. *Medical and Pediatric Oncology*. 2002;38(4):229-239. doi:10.1002/mpo.1316

11. Robison LL, Armstrong GT, Boice JD, et al. The Childhood Cancer Survivor Study: A National Cancer Institute–Supported Resource for Outcome and Intervention Research. *J Clin Oncol*. 2009;27(14):2308-2318. doi:10.1200/JCO.2009.22.3339

12. Leisenring WM, Mertens AC, Armstrong GT, et al. Pediatric Cancer Survivorship Research: Experience of the Childhood Cancer Survivor Study. *Journal of Clinical Oncology*. 2009;27(14):2319. doi:10.1200/JCO.2008.21.1813

13. Armstrong GT, Chen Y, Yasui Y, et al. Reduction in Late Mortality among 5-Year Survivors of Childhood Cancer. *New England Journal of Medicine*. 2016;374(9):833-842. doi:10.1056/NEJMoa1510795

14. Overall CCSS Cohort Demographic and Treatment Exposure Tables. https://ccss.stjude.org/content/dam/en_US/shared/ccss/documents/data/treatment-exposure-tables.pdf. Accessed December 5, 2020.

15. *Childhood Cancer Survivor Study - 2013 Progress Report*. Childhood Cancer Survivor Study; 2013. https://ccss.stjude.org/content/dam/en_US/shared/ccss/documents/progress-report-2013.pdf. Accessed September 20, 2020.

16. Neglia JP, Friedman DL, Yasui Y, et al. Second Malignant Neoplasms in Five-Year Survivors of Childhood Cancer: Childhood Cancer Survivor Study. *J Natl Cancer Inst*. 2001;93(8):618-629. doi:10.1093/jnci/93.8.618

17. Mostoufi-Moab S, Seidel K, Leisenring WM, et al. Endocrine Abnormalities in Aging Survivors of Childhood Cancer: A Report From the Childhood Cancer Survivor Study. *J Clin Oncol*. 2016;34(27):3240-3247. doi:10.1200/JCO.2016.66.6545

18. Green DM, Nolan VG, Goodman PJ, et al. The cyclophosphamide equivalent dose as an approach for quantifying alkylating agent exposure: A report from the childhood cancer survivor study: Cyclophosphamide Equivalent Dose. *Pediatric Blood & Cancer*. 2014;61(1):53-67. doi:10.1002/pbc.24679

19. Sklar CA, Mertens AC, Mitby P, et al. Premature Menopause in Survivors of Childhood Cancer: A Report From the Childhood Cancer Survivor Study. *J Natl Cancer Inst*. 2006;98(13):890-896. doi:10.1093/jnci/djj243

20. Bedoschi G, Navarro PA, Oktay K. Chemotherapy-induced damage to ovary: mechanisms and clinical impact. *Future Oncol*. 2016;12(20):2333-2344. doi:10.2217/fon-2016-0176

21. Kim H, Kim H, Ku S-Y. Fertility preservation in pediatric and young adult female cancer patients. *Ann Pediatr Endocrinol Metab*. 2018;23(2):70-74. doi:10.6065/apem.2018.23.2.70

22. Sonmezer M. Fertility preservation in female patients. *Human Reproduction Update*. 2004;10(3):251-266. doi:10.1093/humupd/dmh021

23. Medrano JV, Andrés M del M, García S, et al. Basic and Clinical Approaches for Fertility Preservation and Restoration in Cancer Patients. *Trends in Biotechnology*. 2018;36(2):199-215. doi:10.1016/j.tibtech.2017.10.010

24. Whb W, Cjh K. Fertility in Female Childhood Cancer Survivors. 2009:24.

25. Wallace WHB, Anderson RA, Irvine DS. Fertility preservation for young patients with cancer: who is at risk and what can be offered? 2005:10.

26. Bath LE, Critchley HOD, Chambers SE, Anderson RA, Kelnar CJH, Wallace WHB. Ovarian and uterine characteristics after total body irradiation in childhood and adolescence: response to sex steroid replacement. *BJOG: An International Journal of Obstetrics & Gynaecology*. 1999;106(12):1265-1272. doi:10.1111/j.1471-0528.1999.tb08180.x

27. Meirow D, Nugent D. The effects of radiotherapy and chemotherapy on female reproduction. *Human reproduction update*. 2001;7(6):535-543.

28. Rodriguez-Wallberg KA, Oktay K. Fertility Preservation Medicine: Options for Young Adults and Children With Cancer. *Journal of Pediatric Hematology/Oncology*. 2010;32(5):390-396. doi:10.1097/MPH.0b013e3181dce339

29. Alkylating antineoplastic agent. In: *Wikipedia*. ; 2020. https://en.wikipedia.org/w/index.php?title=Alkylating_antineoplastic_agent&oldid=962430170. Accessed October 3, 2020.

30. Oktem O, Oktay K. Quantitative assessment of the impact of chemotherapy on ovarian follicle reserve and stromal function. *Cancer*. 2007;110(10):2222-2229. doi:10.1002/cncr.23071

31. Soleimani R, Heytens E, Darzynkiewicz Z, Oktay K. Mechanisms of chemotherapy-induced human ovarian aging: double strand DNA breaks and microvascular compromise. *Aging (Albany NY)*. 2011;3(8):782-793.

32. Spears N, Lopes F, Stefansdottir A, et al. Ovarian damage from chemotherapy and current approaches to its protection. *Hum Reprod Update*. 2019;25(6):673-693. doi:10.1093/humupd/dmz027

33. Bar J, Davidi O, Goshen Y, Hod M, Yaniv I, Hirsch R. Pregnancy outcome in women treated with doxorubicin for childhood cancer. *American Journal of Obstetrics and Gynecology*. 2003;189(3):853-857. doi:10.1067/S0002-9378(03)00837-8

34. Bedoschi GM, Navarro PA, Oktay KH. Novel insights into the pathophysiology of chemotherapy-induced damage to the ovary. *Panminerva Med*. 2019;61(1):68-75. doi:10.23736/S0031-0808.18.03494-8

35. Dasari S, Bernard Tchounwou P. Cisplatin in cancer therapy: Molecular mechanisms of action. *European Journal of Pharmacology*. 2014;740:364-378. doi:10.1016/j.ejphar.2014.07.025

36. Yuksel A, Bildik G, Senbabaoglu F, et al. The magnitude of gonadotoxicity of chemotherapy drugs on ovarian follicles and granulosa cells varies depending upon the category of the drugs and the type of granulosa cells. *Hum Reprod*. October 2015:dev256. doi:10.1093/humrep/dev256

37. Morgan S, Anderson RA, Gourley C, Wallace WH, Spears N. How do chemotherapeutic agents damage the ovary? *Human Reproduction Update*. 2012;18(5):525-535. doi:10.1093/humupd/dms022

38. How Chemotherapy Drugs Work. https://www.cancer.org/treatment/treatments-and-side-effects/treatment-types/chemotherapy/how-chemotherapy-drugs-work.html. Accessed October 2, 2020.

39. van Dorp W, Mulder RL, Kremer LCM, et al. Recommendations for Premature Ovarian Insufficiency Surveillance for Female Survivors of Childhood, Adolescent, and Young Adult Cancer: A Report From the International Late Effects of Childhood Cancer Guideline Harmonization Group in Collaboration With the PanCareSurFup Consortium. *J Clin Oncol*. 2016;34(28):3440-3450. doi:10.1200/JCO.2015.64.3288

40. Muñoz M, Santaballa A, Seguí MA, et al. SEOM Clinical Guideline of fertility preservation and reproduction in cancer patients (2016). *Clin Transl Oncol*. 2016;18(12):1229-1236. doi:10.1007/s12094-016-1587-9

41. Children's oncology group. *Long-Term Follow-Up Guidelines for Survivors of Childhood, Adolescent, and Young Adult Cancers*.; 2018. http://www.survivorshipguidelines.org/.

42. De Vos M, Devroey P, Fauser BC. Primary ovarian insufficiency. *The Lancet*. 2010;376(9744):911-921. doi:10.1016/S0140-6736(10)60355-8

43. Critchley HOD, Wallace WHB. Impact of cancer treatment on uterine function. *J Natl Cancer Inst Monographs*. 2005;(34):64-68. doi:10.1093/jncimonographs/lgi022

44. Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. Third edition. Hoboken, NJ: Wiley; 2019.

45. White IR, Carlin JB. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *Statistics in Medicine*. 2010;29(28):2920-2931. doi:10.1002/sim.3944

46. Rubin DB. Multiple imputations in sample surveys-a phenomenological Bayesian approach to nonresponse. In: Vol 1. American Statistical Association; 1978:20-34.

47. Rubin DB. Multiple Imputation after 18+ Years. *Journal of the American Statistical Association*. 1996;91(434):473-489. doi:10.1080/01621459.1996.10476908

48. Rubin DB, Schenker N. Multiple imputation in health-are databases: An overview and some applications. *Statistics in Medicine*. 1991;10(4):585-598. doi:10.1002/sim.4780100410

49. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Vol 81. John Wiley & Sons; 2004.

50. Schafer JL. *Analysis of Incomplete Multivariate Data*. CRC press; 1997.

51. Buuren SV, Brand JPL, Groothuis-Oudshoorn CGM, Rubin DB. Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation*. 2006;76(12):1049-1064. doi:10.1080/10629360600810434

52. Buuren S van, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*. 2011;45(1):1-67. doi:10.18637/jss.v045.i03

53. van Buuren S, Greenacre M. *Flexible Imputation of Missing Data*. Second.; 2018.

54. Royston P. Multiple imputation of missing values. Stata Journal, 4 (3), 227-241. 2004.

55. Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention science*. 2007;8(3):206-213.

56. Bodner TE. What improves with increased missing data imputations? *Structural Equation Modeling: A Multidisciplinary Journal*. 2008;15(4):651-675.

57. Von Hippel PT. How many imputations do you need? A two-stage calculation using a quadratic rule. *Sociological Methods & Research*. 2020;49(3):699-718.

58. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Statistics in medicine*. 2011;30(4):377-399.

59. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*. 2006;7(1):91. doi:10.1186/1471-2105-7-91

60. Zheng Y, Cai T, Feng Z. Application of the Time-Dependent ROC Curves for Prognostic Accuracy with Multiple Biomarkers. *Biometrics*. 2006;62(1):279-287. doi:10.1111/j.1541-0420.2005.00441.x

61. Robins JM, Rotnitzky A. Semiparametric Efficiency in Multivariate Regression Models with Missing Data. *Journal of the American Statistical Association*. 1995;90(429):122-129. doi:10.1080/01621459.1995.10476494

62. Vock DM, Wolfson J, Bandyopadhyay S, et al. Adapting machine learning techniques to censored time-to-event health record data: A general-purpose approach using inverse probability of censoring weighting. *Journal of Biomedical Informatics*. 2016;61:119-131. doi:10.1016/j.jbi.2016.03.009

63. Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2005;67(2):301-320. doi:10.1111/j.1467-9868.2005.00503.x

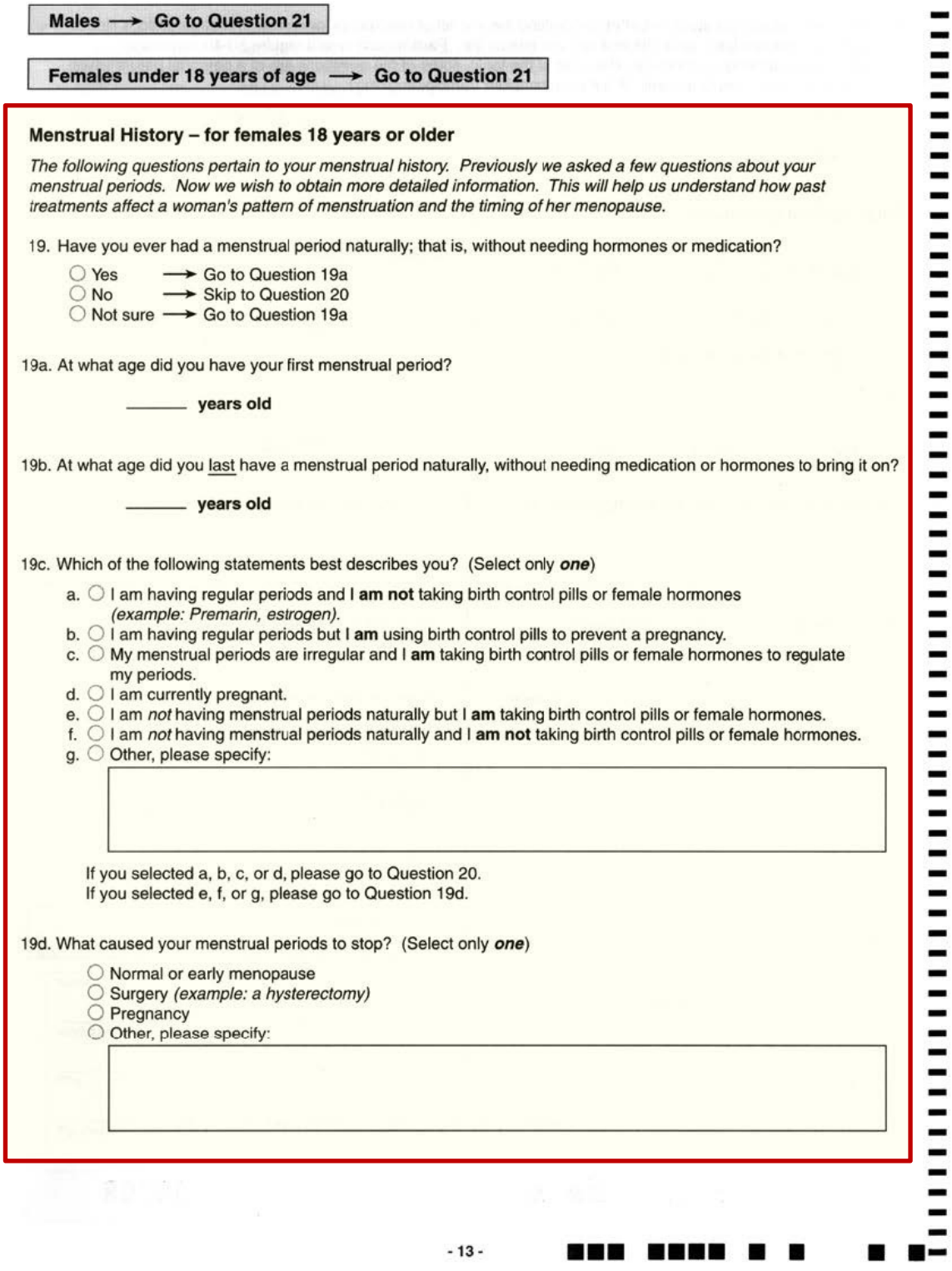
64. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*. San Francisco, California, USA: ACM Press; 2016:785-794. doi:10.1145/2939672.2939785

65. Bergstra J, Bengio Y. Random Search for Hyper-Parameter Optimization. :25.

Appendices

Appendix A Menstrual history survey questions

Ovarian status classifications for female original cohort participants were derived using information from follow-up 1 (items 19-19d), follow-up 4 (items F13-F16, J33-J34), and the follow-up 5 questionnaire (items G13-G16, J35-J36). Ovarian status classifications for female expansion cohort participants were derived using information from the expansion baseline questionnaire (items E13-E16, I33-I34) and the follow-up 5 questionnaire. The surveys administered to the CCSS participants are available at the Childhood Cancer Survivor Study website (https://ccss.stjude.org/tools-and-documents/questionnaires/baseline-and-follow-up-questionnaires.html). Specific questions about ovarian status classification are included below.



Males → Go to Question 21

Females under 18 years of age → Go to Question 21

Menstrual History – for females 18 years or older

The following questions pertain to your menstrual history. Previously we asked a few questions about your menstrual periods. Now we wish to obtain more detailed information. This will help us understand how past treatments affect a woman's pattern of menstruation and the timing of her menopause.

19. Have you ever had a menstrual period naturally; that is, without needing hormones or medication?

- ○ Yes → Go to Question 19a
- ○ No → Skip to Question 20
- ○ Not sure → Go to Question 19a

19a. At what age did you have your first menstrual period?

_____ **years old**

19b. At what age did you last have a menstrual period naturally, without needing medication or hormones to bring it on?

_____ **years old**

19c. Which of the following statements best describes you? (Select only ***one***)

- a. ○ I am having regular periods and I **am not** taking birth control pills or female hormones *(example: Premarin, estrogen).*
- b. ○ I am having regular periods but I **am** using birth control pills to prevent a pregnancy.
- c. ○ My menstrual periods are irregular and I **am** taking birth control pills or female hormones to regulate my periods.
- d. ○ I am currently pregnant.
- e. ○ I am *not* having menstrual periods naturally but I **am** taking birth control pills or female hormones.
- f. ○ I am *not* having menstrual periods naturally and I **am not** taking birth control pills or female hormones.
- g. ○ Other, please specify:

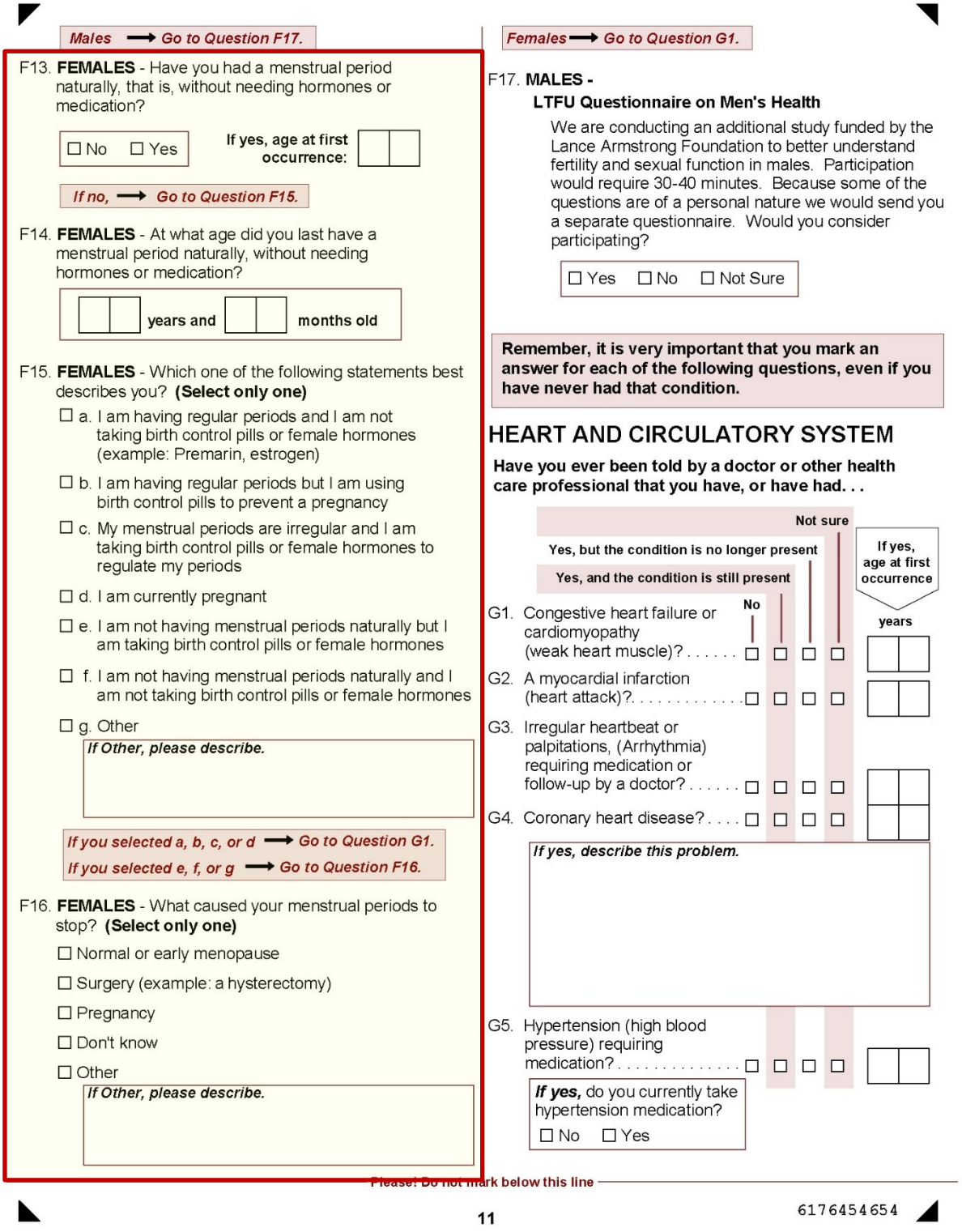
If you selected a, b, c, or d, please go to Question 20.
If you selected e, f, or g, please go to Question 19d.

19d. What caused your menstrual periods to stop? (Select only ***one***)

- ○ Normal or early menopause
- ○ Surgery *(example: a hysterectomy)*
- ○ Pregnancy
- ○ Other, please specify:

- 13 -

Figure A.1 Follow-up 1 survey (2000) Questions 19 and 19a-d



Males → *Go to Question F17.*

F13. **FEMALES** - Have you had a menstrual period naturally, that is, without needing hormones or medication?

☐ No ☐ Yes

If yes, age at first occurrence: ☐☐

If no, → *Go to Question F15.*

F14. **FEMALES** - At what age did you last have a menstrual period naturally, without needing hormones or medication?

☐☐ **years and** ☐☐ **months old**

F15. **FEMALES** - Which one of the following statements best describes you? **(Select only one)**

☐ a. I am having regular periods and I am not taking birth control pills or female hormones (example: Premarin, estrogen)

☐ b. I am having regular periods but I am using birth control pills to prevent a pregnancy

☐ c. My menstrual periods are irregular and I am taking birth control pills or female hormones to regulate my periods

☐ d. I am currently pregnant

☐ e. I am not having menstrual periods naturally but I am taking birth control pills or female hormones

☐ f. I am not having menstrual periods naturally and I am not taking birth control pills or female hormones

☐ g. Other

If Other, please describe.

If you selected a, b, c, or d → *Go to Question G1.*
If you selected e, f, or g → *Go to Question F16.*

F16. **FEMALES** - What caused your menstrual periods to stop? **(Select only one)**

☐ Normal or early menopause

☐ Surgery (example: a hysterectomy)

☐ Pregnancy

☐ Don't know

☐ Other

If Other, please describe.

Females → *Go to Question G1.*

F17. **MALES -**

LTFU Questionnaire on Men's Health

We are conducting an additional study funded by the Lance Armstrong Foundation to better understand fertility and sexual function in males. Participation would require 30-40 minutes. Because some of the questions are of a personal nature we would send you a separate questionnaire. Would you consider participating?

☐ Yes ☐ No ☐ Not Sure

Remember, it is very important that you mark an answer for each of the following questions, even if you have never had that condition.

HEART AND CIRCULATORY SYSTEM

Have you ever been told by a doctor or other health care professional that you have, or have had. . .

	No	Yes, but the condition is no longer present	Yes, and the condition is still present	Not sure	If yes, age at first occurrence
G1. Congestive heart failure or cardiomyopathy (weak heart muscle)?	☐	☐	☐	☐	☐☐ years
G2. A myocardial infarction (heart attack)?	☐	☐	☐	☐	☐☐
G3. Irregular heartbeat or palpitations, (Arrhythmia) requiring medication or follow-up by a doctor?	☐	☐	☐	☐	☐☐
G4. Coronary heart disease?	☐	☐	☐	☐	☐☐

If yes, describe this problem.

G5. Hypertension (high blood pressure) requiring medication?	☐	☐	☐	☐	☐☐

If yes, do you currently take hypertension medication?
☐ No ☐ Yes

Please! Do not mark below this line

11

6176454654

Figure A.2 Follow-up 4 survey (2007) Questions F13-16

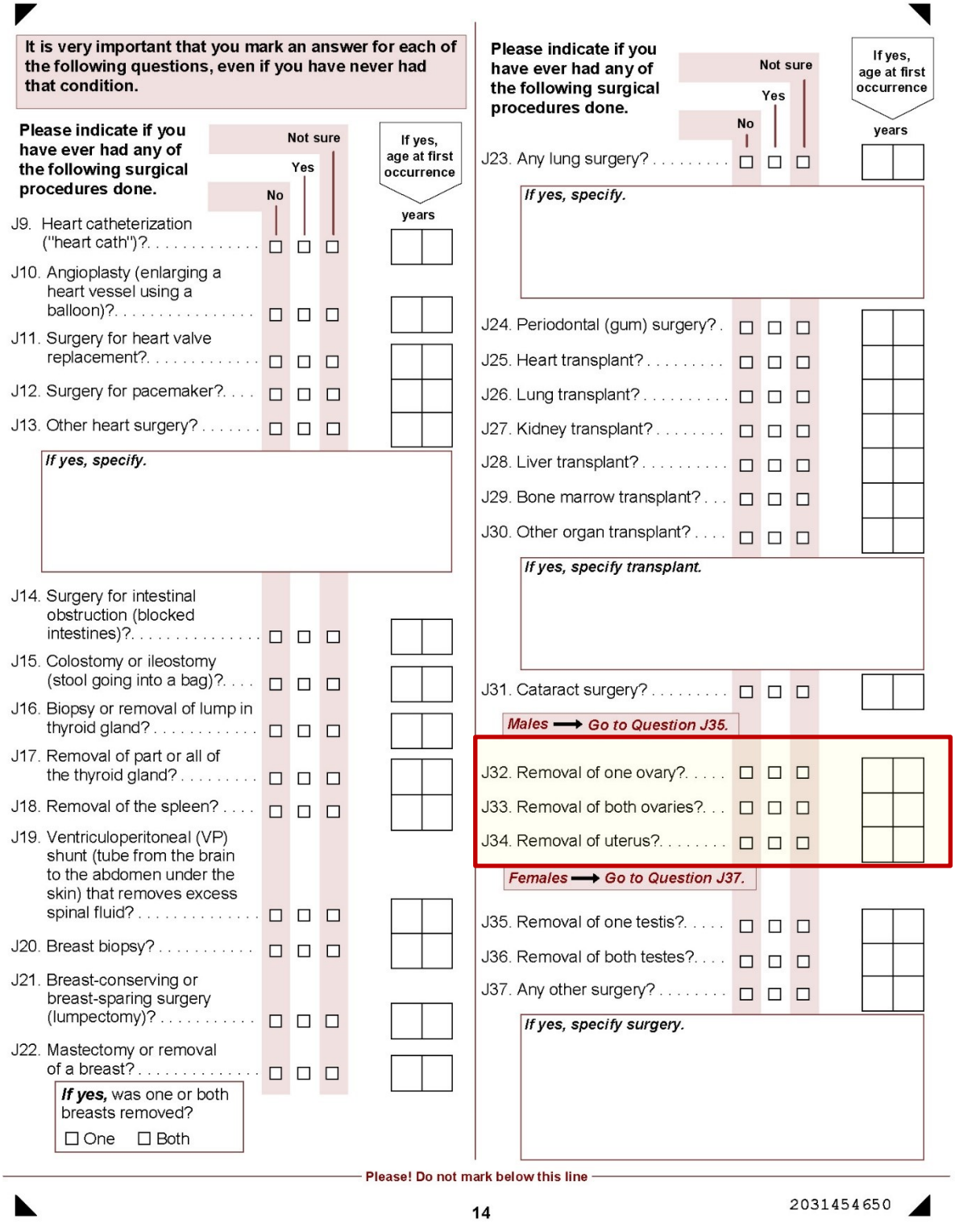


Figure A.3 Follow-up 4 survey (2007) Questions J32-34

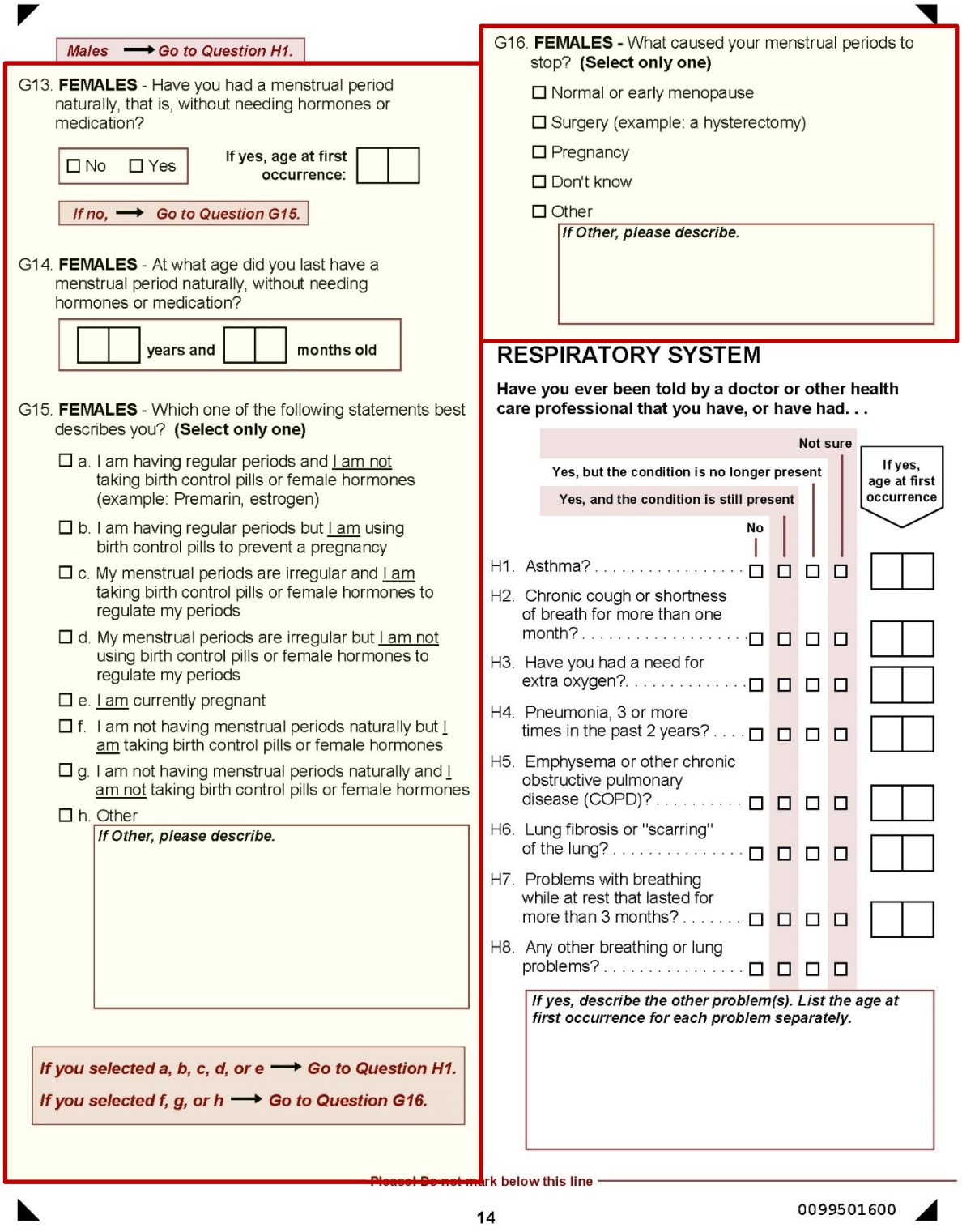


Figure A.4 Follow-up 5 survey (2014) Questions G13-15

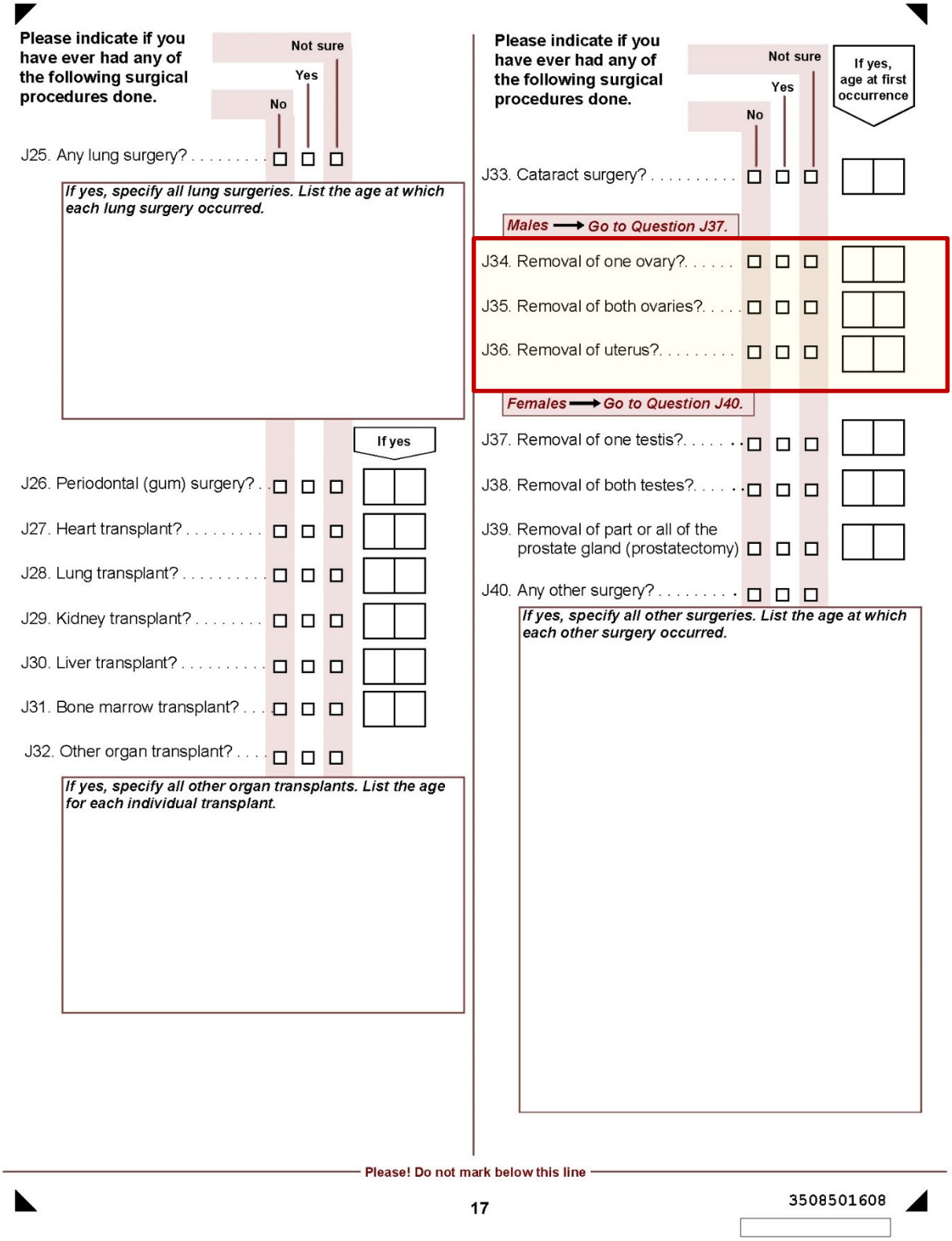


Figure A.5 Follow-up 5 survey (2014) Questions J34-36

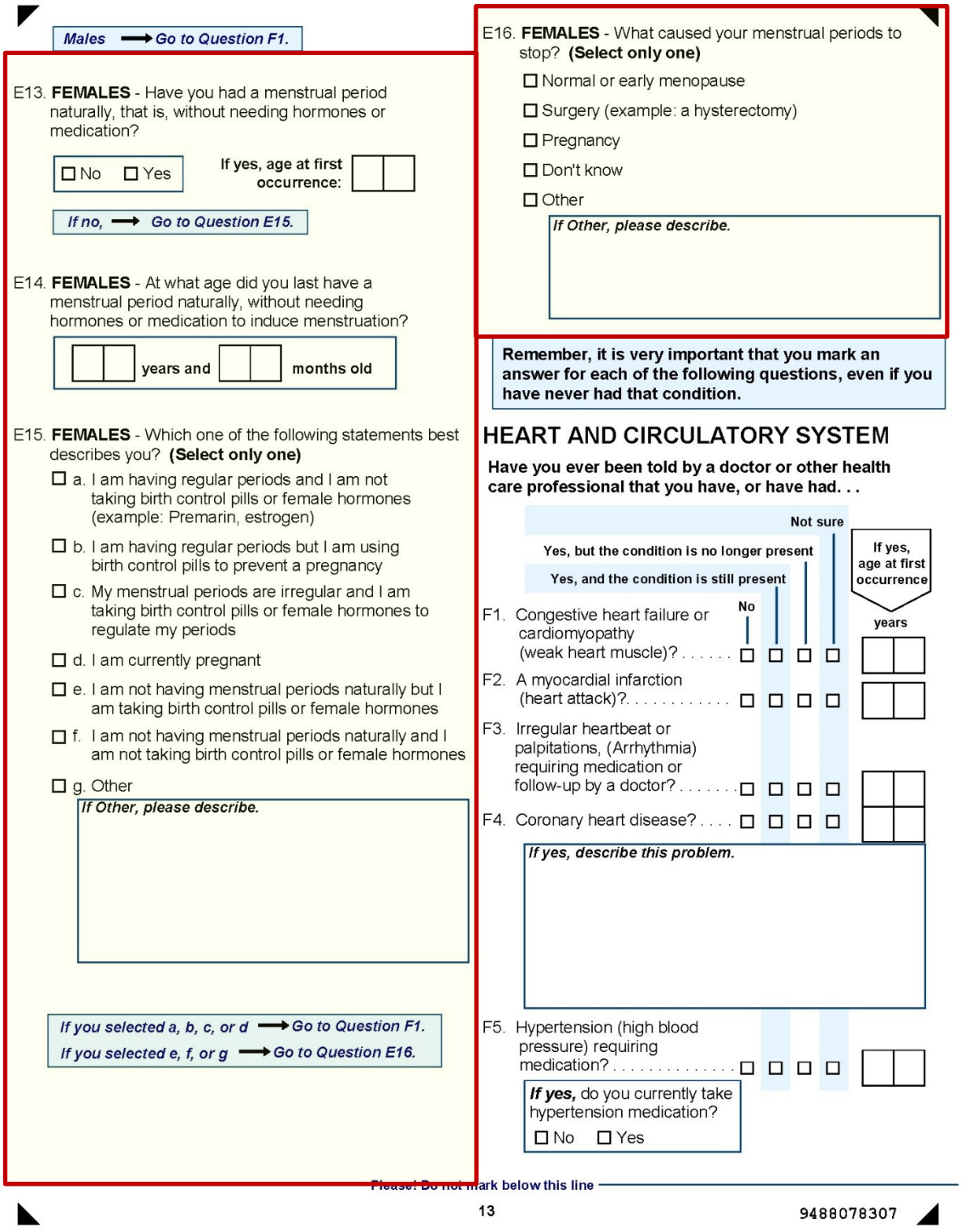


Figure A.6 Expansion cohort baseline survey (2008 – present) Questions E13-16

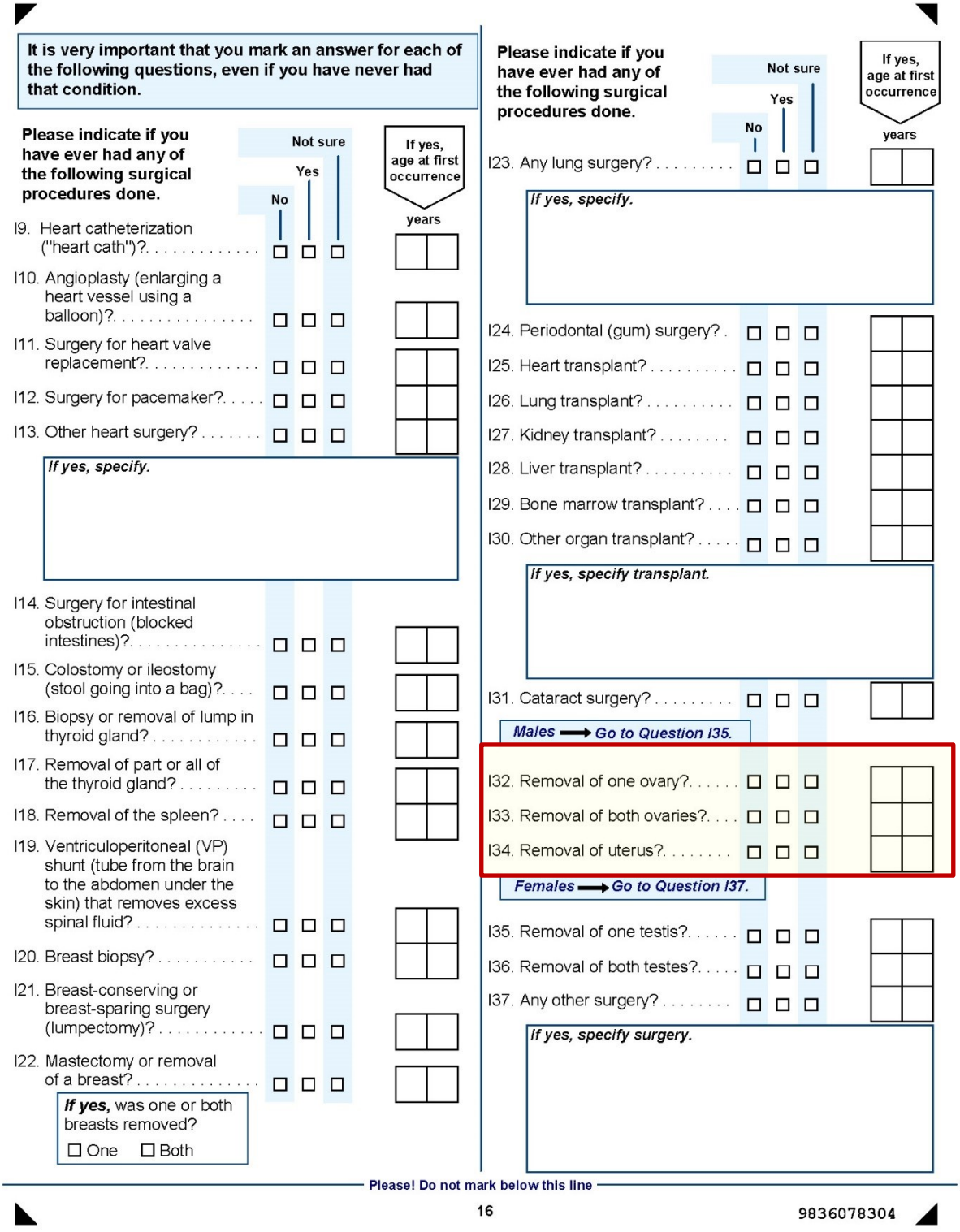


Figure A.7 Expansion cohort baseline survey (2008 – present) Questions I32-34

Appendix B Age at event/cesoring assignment

Variables' description

Table B.1 Variables for assigning age at event/censoring

Variables	Description	Source
a_menslast_overall (some reported years and months)	age at last natural menstrual period	self-reported in expansion baseline, follow-up 1, 4 and 5 surveys
menarcheage (only in years)	menarche age	self-reported in expansion baseline, follow-up 1, 4 and 5 surveys
a_SurgicalPM (only in years)	age at surgical time	self-reported in "SURGICAL PROCEDURES" in expansion baseline, follow-up 4 and 5 surveys
age_dx (exact age)	age at primary cancer diagnosis	calculated based on the date of cancer diagnosed
age_smn1 (exact age)	age at second malignant neoplasm	calculated based on the date of second cancer diagnose
age_lastmhr (exact age)	age at last survey that contained menstrual history	calculated based on the date of the most recent questionnaire contained menstrual history
statusgoli_overall	status in original data set from CCSS	
a_event	new generated age at event	

Algorithms

Table B.2 Algorithms for assignment of age at event / censoring

Status	Status definition	age at event / censoring	Notes / Modification
NSPM	ovarian function retained for at least 5 years following cancer diagnosis, but menopause before age 40	age at last menstrual period	If the age at last menstrual was too young (e.g. age < 9), the record is deemed incorrect. The subjects will be excluded from analysis.
Surgical PM	Surgery was the reason cited for the onset of menopause prior age 40	1. age at surgical time	Some "a_menslast_overall" were much earlier than "a_SurgicalPM", they were reclassified as *NSPM or AOF* and age at event = "*a_menslast_overall*"
		2. age at last menstrual period	"a_menslast_overall" might occur much earlier than "a_SurgicalPM", in which case, they should be NSPM or AOF. To avoid misclassifying the outcome, they were excluded depending on the cut-off age for outcomes.
AOF	Patient's menstrual period stopped permanently within 5 years of cancer diagnosis	menarche age or age at last menstrual period or 16, or a_dx + 5	If a_menslast_overall is missing, we use age at menarche as the age at event. If both are missing, we use the maximum of (16, age at diagnosis plus 5).
AOF non menarche	Patients reported never going through menarche by age 18	a_dx + 5 or 16	There is no menstrual history to assign age at event for this group. The maximum of (age at diagnose + 5, 16) is used.
Normal	No indications that periods ceased prior to age 40 due to any of the above causes	1. age at last survey with menstrual history or 40	Extensions of menstrual status beyond age 40 would require further review of the available data.
		2. age at last menstrual period	When this age was greater than 40.
		3. age at SMN	When SMN prior to age at last menstrual period, the age at event is the age of SMN with a normal ovarian status.

Changed status (N = 102)

Table B.3 Summary of changed status

Original status	New status	N = 102	Reasons
Surgical PM	Normal	27	SMN prior to surgical time.
		1	Age at surgical time and age at the last were greater than 40, therefore she had normal ovarian status before 40 and was assigned to normal group.
	AOF	7	Last menstrual period occurs much earlier than surgery and the time are within 5 years of cancer diagnosis.
	NSPM	37	Last menstrual period occurs much earlier than surgery at an age after 5 years of cancer diagnosis.
NSPM	Normal	29	NSPM occurs before surgical time.
PM (possible)	Normal	1	SMN happens prior to PM

Table B.4 Excluded subjects because of conflicting age information (N = 27 + 129)

Original status	Reasons	Numbers
Normal	age at event < 9 years old	1
NSPM	age at event < 9 years old	1
Surgical PM	age at event < 9 years old	17
Surgical PM	age at event occurred 3 years prior to age at diagnosis	3
AOF	age at event occurred 3 years prior to age at diagnosis	5
Surgical PM	age at surgical time were missing	129

Appendix C Data dictionary of final work data set

Table C.1 Dictionary of final data set (62 variables)

	Variable name	Data type	Levels / Range (units)	Description
ID	ccssid			CCSSI ID
INFO (3 vars)	cohort	categorical	2 levels: original, expansion	which cohort the patient was recruited in
	d_birth	date	[1949-09-12, 1998-08-25]	date of birth
	race_3	race	3 levels: white, black, other	race of survivors
Outcome (2 vars)	status	categorical	3 levels: Normal, POI, SPM	status of ovarian function
	a_event	numerical	[9.9, 57.0] (years)	age at event
diagnose (2 vars)	diagnose	categorical	8 levels: Leukemia, CNS, HD, HNL, Kidney (Wilms), Neuroblastoma, Soft tissue sarcoma, Bone cancer	Cancer diagnose type
	age_dx	numerical	[0-21) (years)	age at diagnose
BMT	bmt_tbi	categorical	Yes/No	Bone marrow transplant indicator
Radiotherapy (within 5 years of first cancer) （5 body regions）(10 vars)	tbidose	numerical	[0, 1575] (cGy)	total body irradiation dose
	abdmaxrtdose	numerical	[0, 6900] (cGy)	cumulative radiation doses to abdomen
	pelvismaxrtdose	numerical	[0, 7800] (cGy)	cumulative radiation doses to pelvis
	pitdose	numerical	[0, 2940] (cGy)	cumulative radiation doses to pituitary
	minovary	numerical	[0, 5940] (cGy)	minimum cumulative radiation dose to ovary
	maxovary	numerical	[0, 7800] (cGy)	maximum cumulative radiation dose to ovary
	rt_yn	categorical	Yes/No	whether the patient received radio therapy
	tbirt_yn	categorical	Yes/No	whether the patient received total body irradiation
	abdomenrt_yn / pelvisrt_yn	categorical	Yes/No	whether the patient received irradiation to abdomen / pelvis

	Variable name	Data type	Levels / Range (units)	Description
Chemotherapy (within 5 years of first cancer) (21 chemotherapy agents and CED) (43 vars)	ced5		[0, 83301] (mg/m^2)	Cyclophosphamide Equivalent Dose[1]
	bcnu		[0, 1573] (mg/m^2)	cumulative dose of BCNU (Alkylating agent)
	busulfan		[0, 650] (mg/m^2)	cumulative dose of busulfan (Alkylating agent)
	ccnu		[0, 1333] (mg/m^2)	cumulative dose of CCNU (Alkylating agent)
	chlorambucil		[0, 3349] (mg/m^2)	cumulative dose of chlorambucil (Alkylating agent)
	cyclophosphamide		[0, 83301] (mg/m^2)	cumulative dose of cyclophosphamide (Alkylating agent)
	ifosfamide		[0, 144230] (mg/m^2)	cumulative dose of ifosfamide (Alkylating agent)
	melphalan		[0, 514] (mg/m^2)	cumulative dose of melphalan (Alkylating agent)
	nitrogen_mustard		[0, 256] (mg/m^2)	cumulative dose of nitrogen_mustard (Alkylating agent)
	procarbazine		[0, 17500] (mg/m^2)	cumulative dose of procarbazine (Alkylating agent)
	thiotepa		[0, 933] (mg/m^2)	cumulative dose of thiotepa (Alkylating agent)
	carboplatin		[0, 15711] (mg/m^2)	cumulative dose of carboplatin (Platinum compounds)
	cis_platinum		[0, 7075] (mg/m^2)	cumulative dose of cis_platinum (Platinum compounds)
	bleomycin		[0, 402] (mg/m^2)	cumulative dose of bleomycin (Antibiotics)
	daunorubicin		[0, 838] (mg/m^2)	cumulative dose of daunorubicin (Anthracyclines)
	doxorubicin		[0, 1070] (mg/m^2)	cumulative dose of doxorubicin (Anthracyclines)
	epirubicin		[0, 0] (mg/m^2)	cumulative dose of epirubicin (Anthracyclines)
	idarubicin		[0, 192] (mg/m^2)	cumulative dose of idarubicin (Anthracyclines)
	methotrexate		[0, 502553] (mg/m^2)	cumulative dose of methotrexate (Anthracyclines)
	mitoxantrone		[0, 97] (mg/m^2)	cumulative dose of mitoxantrone (Anthraquinone)
	vm_26		[0, 9300] (mg/m^2)	cumulative dose of vm_26 (Epipodophyllotoxins)
	vp_16		[0, 20594] (mg/m^2)	cumulative dose of vp_16 (Epipodophyllotoxins)
	bcnu_yn, and other 20 agents	categorical	Yes/No	whether the patient received the chemotherapy agent

[1] Daniel M. Green et al., "The Cyclophosphamide Equivalent Dose as an Approach for Quantifying Alkylating Agent Exposure: A Report from the Childhood Cancer Survivor Study: Cyclophosphamide Equivalent Dose," *Pediatric Blood & Cancer* 61, no. 1 (January 2014): 53–67, https://doi.org/10.1002/pbc.24679.

Appendix D Components by ovarian status at different cut-off ages

Table D.1 Components by ovarian status at different cut-off ages

Cut-off age years	Normal n (%)	SPM n (%)	POI n (%)	Censoring n (%)	Total n
21	6809 (86.4)	13 (0.2)	607 (7.7)	416 (5.3)	7883
22	6605 (83.8)	16 (0.2)	639 (8.1)	582 (7.4)	7880
23	6417 (81.5)	18 (0.2)	658 (8.4)	747 (9.5)	7878
24	6208 (78.8)	19 (0.2)	680 (8.6)	931 (11.8)	7876
25	5968 (75.8)	23 (0.3)	703 (8.9)	1140 (14.5)	7872
26	5707 (72.6)	28 (0.4)	718 (9.1)	1374 (17.5)	7865
27	5475 (69.7)	40 (0.5)	727 (9.3)	1575 (20.1)	7855
28	5201 (66.3)	47 (0.6)	736 (9.4)	1824 (23.2)	7846
29	4925 (62.8)	60 (0.8)	748 (9.5)	2072 (26.4)	7843
30	4617 (58.9)	72 (0.9)	756 (9.7)	2350 (30)	7833
31	4296 (54.9)	91 (1.2)	782 (10)	2619 (33.5)	7826
32	3987 (51)	111 (1.4)	793 (10.1)	2891 (37)	7820
33	3667 (46.9)	135 (1.7)	813 (10.4)	3162 (40.5)	7815
34	3372 (43.2)	148 (1.9)	829 (10.6)	3422 (43.8)	7809
35	3034 (38.9)	173 (2.2)	841 (10.8)	3716 (47.6)	7802
36	2715 (34.9)	197 (2.5)	864 (11.1)	3975 (51)	7789
37	2424 (31.1)	220 (2.8)	878 (11.3)	4223 (54.3)	7783
38	2127 (27.4)	234 (3)	888 (11.4)	4488 (57.7)	7775
39	1894 (24.4)	258 (3.3)	905 (11.7)	4668 (60.1)	7763
40	1682 (21.7)	273 (3.5)	917 (11.8)	4852 (62.5)	7762

Appendix E Nested cross-validation

The nested cross-validation includes two loops: the inner CV and outer CV. They serve different purposes: the inner CV is used to simulate traditional CV which is used for tuning model parameters, and the outer CV (the remaining untrained fold) is used for evaluating model performance. The aim of nested CV is to separate the parameter tuning process and model evaluation process. And different from traditional evaluation methods, the nested CV focuses on assessing the modeling procedure rather than the model itself. Therefore, in this step, we do not evaluate a model with specific hyper-parameters, instead, we obtained a validated performance for traditional CV. And in the modeling process, traditional CV was used to find optimal hyper-parameters for fitting in the whole sample data to get final models,

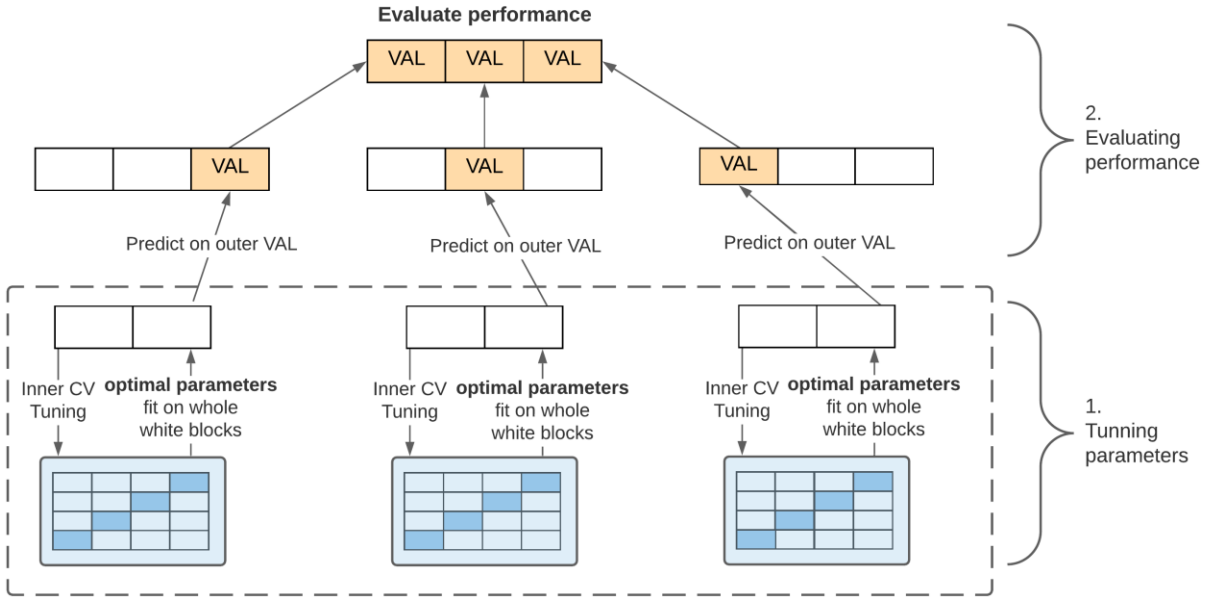


Figure E.1 Procedure of nested CV.
(VAL: validation data sets)

Figure E.1 illustrates the process of nested cross-validation. For better presenting the process in this figure, the outer folds were simplified to three folds instead of five folds in this study, and inner CV was simplified to four-folds.

The process is described below:

Step 1: the data set is split into three folds, each fold will serve as an outer validation set once.

Step 2: the two training folds are split into four folds to perform an inner CV for tuning hyper-parameters.

Step3: the optimal hyper-parameter setting is then used to fit the two white folds (training data set) and predict on the yellow outer fold (validation data set).

Step4: Repeat this procedure three times for each yellow fold and then combine the predicted risk across the whole data to generate a validated performance.

Appendix F　Assessment of convergence in MI

The number of the iteration times was set to 30 to examine the convergence. Figure F.1， Figure F.2, and Figure F.3 visualized the convergence of imputed values for all the variables by plotting the means and standard deviation (SD). There is no convergence issue for any variables. And they attain convergency by 10 iterations.

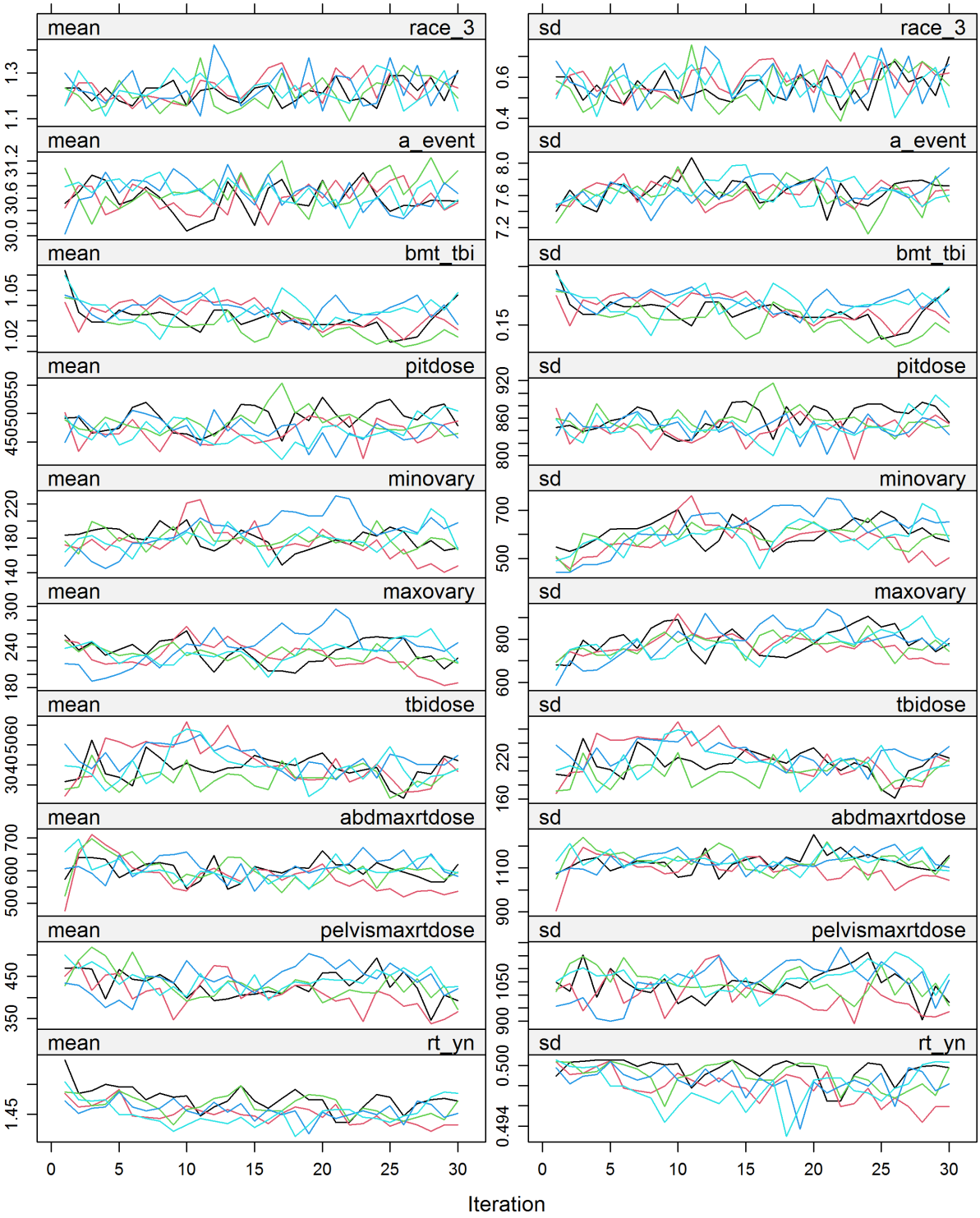


Figure F.1 Convergence plots for variables: race, age at event, BMT indicator, irradiation dose to pituitary, minimum irradiation dose to ovaries, maximum irradiation dose to ovaries, total body irradiation dose, irradiation dose to abdomen, irradiation dose to pelvis, indicator for receiving radiotherapy

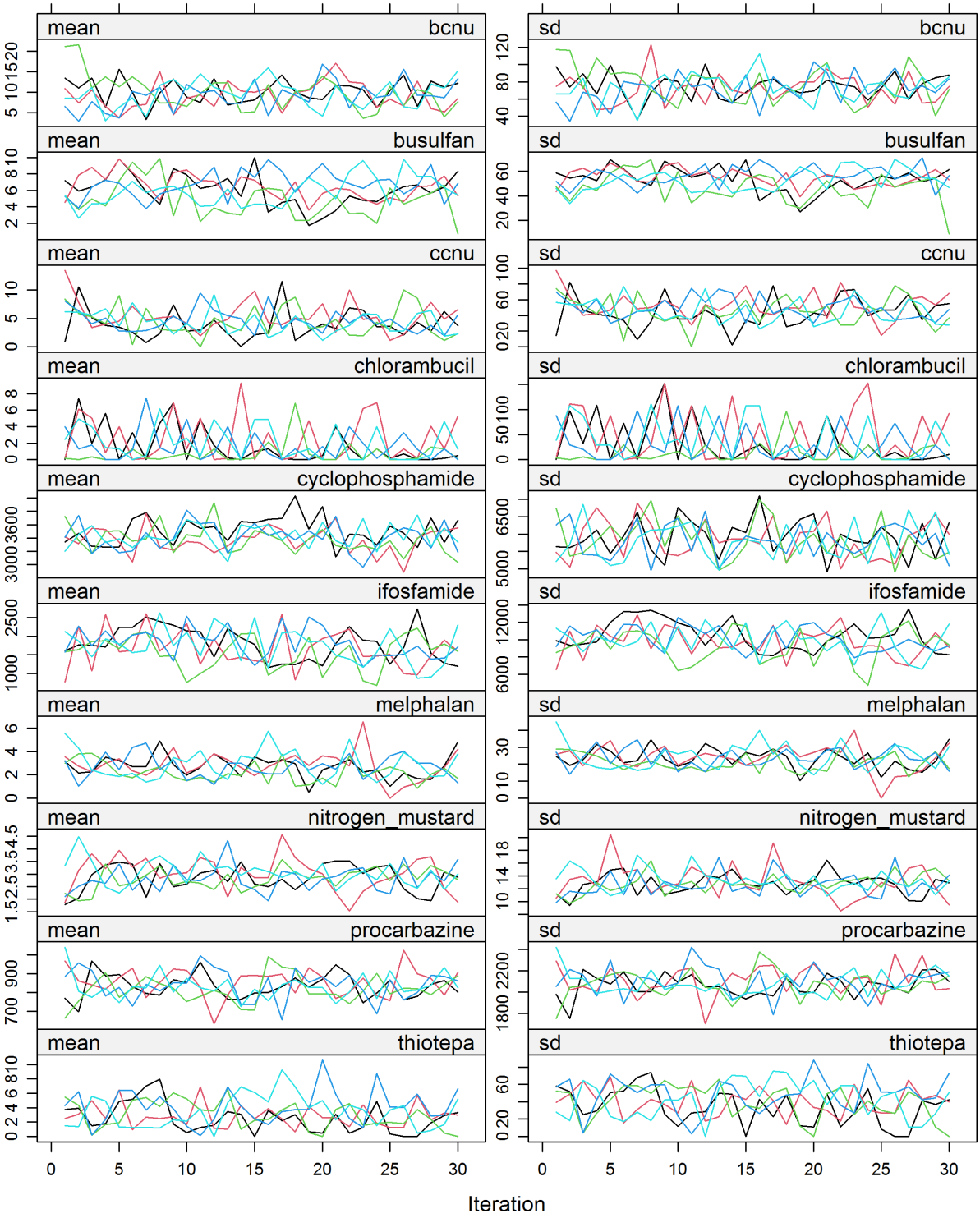


Figure F.2 Convergence plots for doses of 10 chemotherapy agents: bcnu, busulfan, ccnu, chlorambucil, cyclophosphamide, Ifosfamide, melphalan, nitrogen mustard, procarbazine, thiotepa

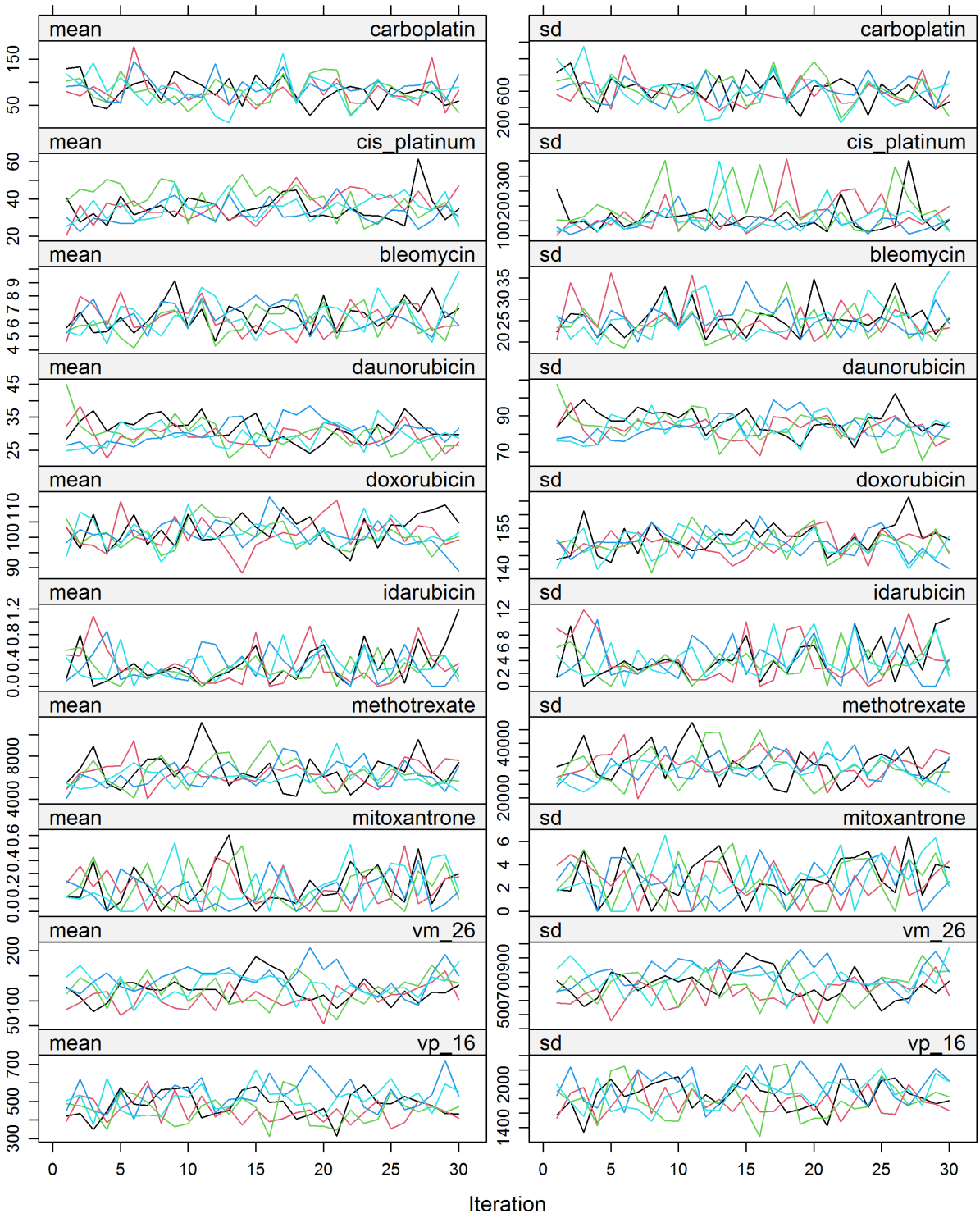


Figure F.3 Convergence plots for doses of 10 chemotherapy agents: carboplatin, cis platinum, bleomycin, daunorubicin, doxorubicin, idarubcin, methotrexate, mitoxantrone, vm_26, vp_16

Appendix G Modeling algorithms

EN-ALR:

Elastic Net is a regularizations method by combining the penalty from LASSO (ℓ_1 penalty) and Ridge (ℓ_2 penalty) regression, which aims to avoid overfitting at the cost of increased bias. It enables automatic variable selection (a feature of LASSO) and avoids the limitation of LASSO regression at the same time.[2]

For logistic regression, the objective function for the penalized logistic regression uses the following log-likelihood:

$$\min_{(\beta_0,\beta)\in\mathbb{R}^{p+1}} -\left[\frac{1}{N}\sum_{i=1}^{N} y_i \cdot (\beta_0 + x_i^T\beta) - log\left(1 + e^{(\beta_0 + x_i^T\beta)}\right)\right] + \lambda[(1-\alpha)\|\beta\|_2^2/2 + \alpha\|\beta\|_1],$$

Wherein β_0 and β are coefficients in the generalized linear model, y_i is the binary outcome for the ith individual, x_i is the vector of covariates of the ith individual, $\|\beta\|_1$ is the ℓ_1 penalty on coefficients of x_i, i.e. β, and $\|\beta\|_2^2$ is the ℓ_2 penalty on β. The two hyperparameters: α and λ control the penalty function, wherein α bridges the gap between LASSO (α=1) and Ridge (α=0) and λ controls the overall strength of the penalty.

XGBoost:

XGBoost refers to "Extreme Gradient Boosting", which is a fast implementation of a gradient boosting algorithm that uses a gradient boosting framework[3]. It has been successfully used in many applications and becomes the winning solution for best predictive performance in numerous

2 Hui Zou and Trevor Hastie, "Regularization and Variable Selection via the Elastic Net," Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67, no. 2 (April 2005): 301–20, https://doi.org/10.1111/j.1467-9868.2005.00503.x.

3 Tianqi Chen and Carlos Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16 (the 22nd ACM SIGKDD International Conference, San Francisco, California, USA: ACM Press, 2016), 785–94, https://doi.org/10.1145/2939672.2939785; Avinash Barnwal, Hyunsu Cho, and Toby Dylan Hocking, "Survival Regression with Accelerated Failure Time Model in XGBoost," ArXiv:2006.04920 [Cs, Stat], June 11, 2020, http://arxiv.org/abs/2006.04920.

competitions.4 Hyperparameter tuning is the key to achieving accurate prediction, however, the cost is computation time. Therefore, to balance accuracy and efficiency, three hyperparameters: *max_depth*, *eta*, and *nrounds* of top importance were finely tuned while default values were used for other hyperparameters. The parameter *max_depth* refers to the maximum depth of a tree, increasing this value will result in a more complex model and more likely to overfit. *eta* stands for the step size shrinkage used in the update to prevent overfitting and *nrounds* controls the maximum number of iterations.

Ensemble:

This method combines multiple algorithms to generate a predicted risk with better predictive performance. Typically, the predicted risks were incorporated using weights which can be tuned as well. In this project, the weights for both algorithms (EN-ALR and XGBoost) were set at 0.5.

Hyperparameter tuning:

50 hyperparameter settings for EN-ALR and XGBoost respectively were randomly selected from hyperparameter spaces (EN-ALR: $\alpha \in [0.05, 0.3]$, and $\lambda \in [0.05,0.3]$; XGBoost: $max\_depth \in [5, 30]$, $eta \in [0.1, 0.5]$, and $nrounds \in [10, 150]$) which were obtained from a manually coarse tuning. Then the optimal hyperparameter setting for each algorithm was determined from the 50 settings based on a weighted sum of the AUC, AP, and sBrS.

A weighted sum of AUC, AP, and sBrS:

AUC, AP, and sBrS are the metrics used to evaluate models. To incorporate the three metrics, an equal-weighted sum of the three metrics was used to find the optimal hyperparameters. In addition, to avoid one metric dominate the rank of the weighted sum due to its magnitude in the 50 hyperparameter settings, AUC, AP, and sBrS were scaled to [0, 1] before being weighted, i.e.

4 Didrik Nielsen, "Tree Boosting With XGBoost," n.d., 110.

$$AUC_{scaled} = [AUC - min(AUC)] \times \frac{1}{\max(\text{AUC}) - \min(\text{AUC})}$$

$$AP_{scaled} = [AP - min(AP)] \times \frac{1}{max(AP) - min(AP)}$$

$$sBrS_{scaled} = [sBrS - min(sBrS)] \times \frac{1}{max(sBrS) - min(sBrS)}$$

Then the weighted sum of the three metrics can be expressed as:

$$\frac{1}{3}(AUC_{scaled} + AP_{scaled} + sBrS_{scaled})$$

Modification of predicted risks

It should be noted that the predicted risks do not have a strictly monotone increasing relationship with ages, as the prediction models for different ages were developed separately. To avoid the occasional decrease in predicted risks, we force the predicted risks at age A to be equal to or greater than the maximum of predicted risks at ages ≤ A, i.e.

$$Risk_A^{modified} = \max(\text{risk}_A, \text{risk}_{A\text{-}})$$

risk_A = predicted risk by age A

$\text{risk}_{A\text{-}}$ = predicted risk by ages younger than A

Predictors

Table 4A listed the predictors used in modeling. For EN-ALR, chemotherapy agents that were rarely used in the study sample, such as busulfan, CCNU, chlorambucil, melphalan, thiotepa, idarubicin, mitoxantrone, were coded as binary Yes/No. In contrast to regression methods, XGBoost, as a tree-based machine learning algorithm, "prefers" continuous variables than categorical variables because it can split it at any point that minimizes the loss function. Therefore, doses of chemotherapy agents were used in developing the XGBoost model.

Table G.1 Predictors used in developing EN-ALR and XGBoost algorithms
Difference between predictors in two algorithms were shaded in blue and orange

Variable Description	EN-ALR	XGBoost
Race (3 levels)	Categorical	Categorical
Age at Cancer Diagnosis	Continuous	Continuous
BMT Indicator	Binary	Binary
Cancer Diagnosis Type (8 levels)	Categorical	Categorical
Minimum Ovarian Radiation Dose	Continuous	Continuous
Radiation doses to pituitary	Continuous	Continuous
Total body irradiation dose	Continuous	Continuous
CED	Continuous	Continuous
BCNU	Continuous	Continuous
Busulfan	Binary	Continuous
CCNU	Binary	Continuous
Chlorambucil	Binary	Continuous
Cyclophosphamide	Continuous	Continuous
Ifosfamide	Continuous	Continuous
Melphalan	Binary	Continuous
Nitrogen Mustard	Continuous	Continuous
Procarbazine	Continuous	Continuous
Thiotepa	Binary	Continuous
Carboplatin	Continuous	Continuous
Cis_Platinum	Continuous	Continuous
Bleomycin	Continuous	Continuous
Daunorubicin	Continuous	Continuous
Doxorubicin	Continuous	Continuous
Idarubicin	Binary	Continuous
Methotrexate	Continuous	Continuous
Mitoxantrone	Binary	Continuous
VM 26	Continuous	Continuous
VP 16	Continuous	Continuous
Interaction: Age at Cancer Diagnosis and BMT	Continuous	NA
Interaction: Age at Cancer Diagnosis and Minimum Ovarian RT Dose	Continuous	NA

Appendix H Validated performance of models with CED and models with 10 individual alkylating agents

The three figures in this appendix showed that models with ten individual alkylating agents (red) had superior performance comparing to models with CED (blue) at most of age thresholds.

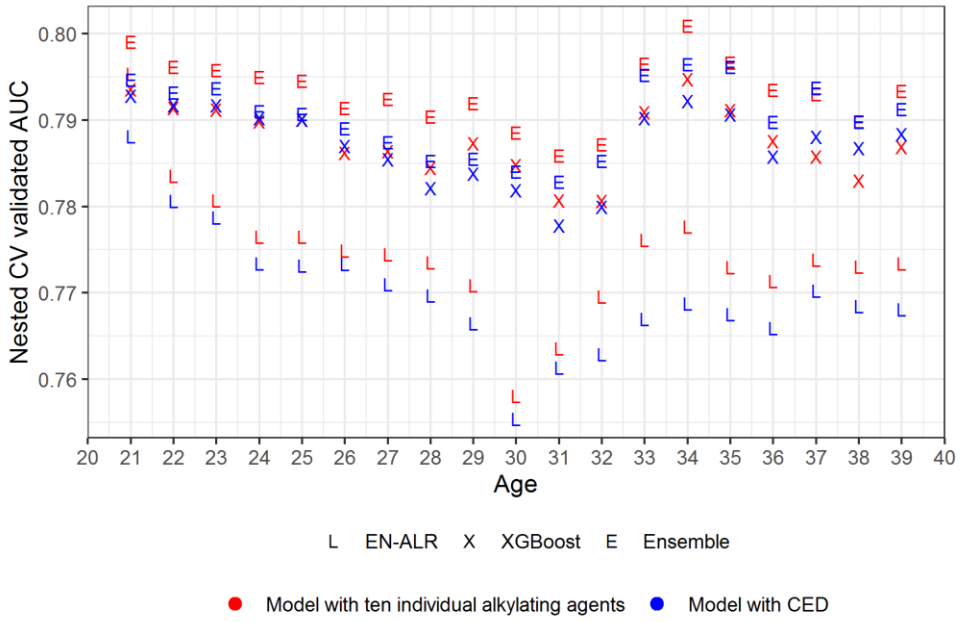


Figure H.1 Nested CV validated AUC between models with 10 individual alkylating agents and models with CED

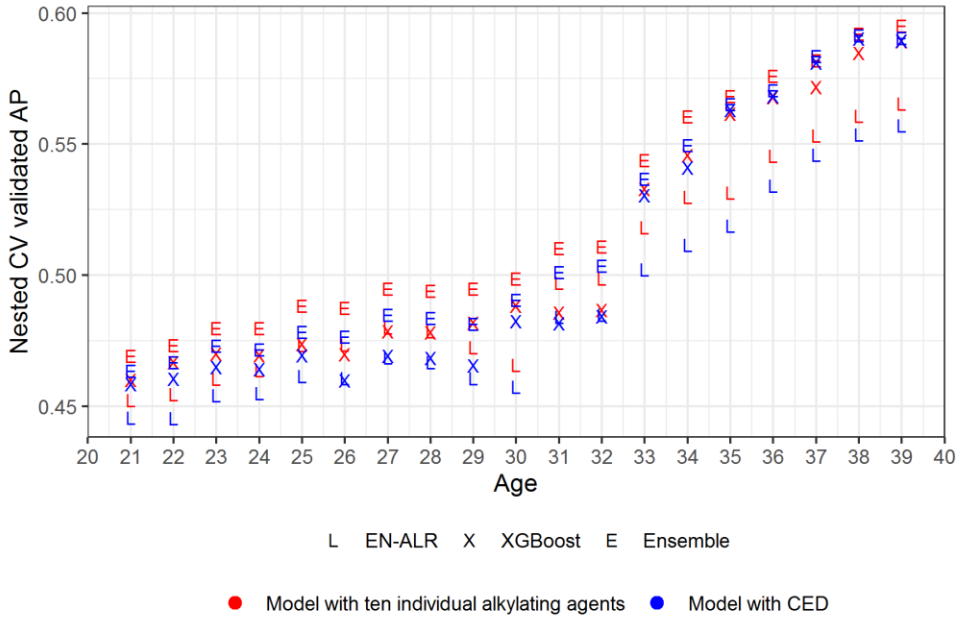


Figure H.2 Nested CV validated AP between models with 10 individual alkylating agents and models with CED

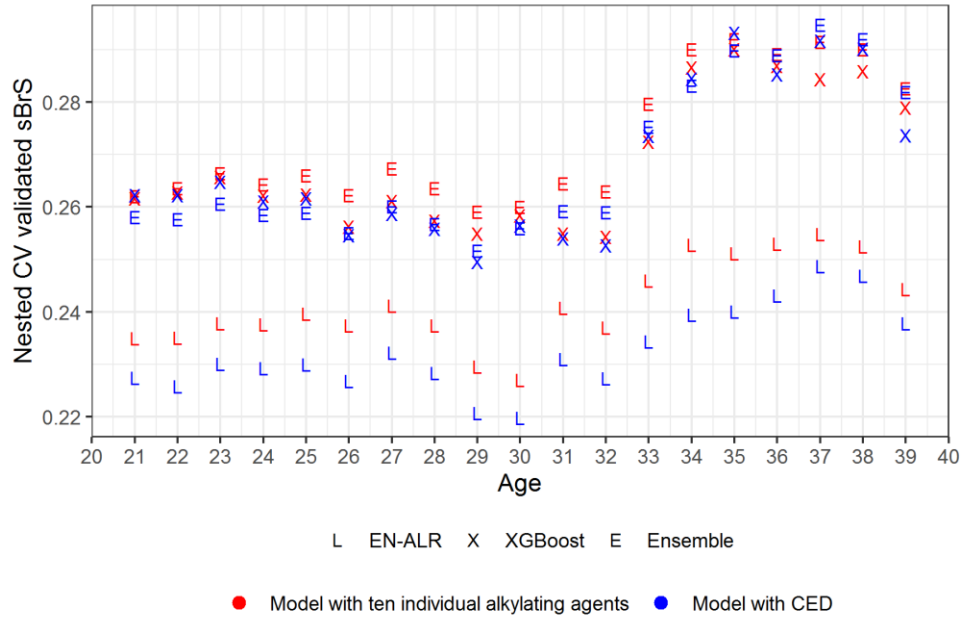


Figure H.3 Nested CV validated sBrS between models with 10 individual alkylating agents and models with CED

Appendix I Application of the prediction tool

The relevant treatment information table is provided for input values. The predicted risks of POI from age 21 to 39 are plotted based on the information in the table. Two examples are presented below.

Example 1 (Table B1 and Figure B1):

Suppose one girl (race: white) was diagnosed with Leukemia at age 2. During the cancer treatment, she received radiotherapy including the irradiation to two ovaries and pituitary. The lower cumulative doses received between both side ovaries was 8.1 cGy and the pituitary received 1800 cGy radiation doses. Her chemotherapy included a total cumulative cyclophosphamide dose 13007 mg/m^2, daunorubicin dose 520 mg/m^2, doxorubicin dose 90 mg/m^2, methotrexate dose 460 mg/m^2, and VM 26 dose 6338 mg/m^2.

Example 2 (Table B2 and Figure B2):

Suppose one girl (race: Asian or Pacific Islander) was diagnosed with Non-Hodgkin lymphoma at age 14. During the cancer treatment, she received radiotherapy including the irradiation to two ovaries and pituitary. The lower cumulative doses received between both side ovaries was 2.3 cGy and the pituitary received 47 cGy radiation doses. Her chemotherapy included a total cumulative BCNU 300 mg/m^2, cyclophosphamide dose 15576 mg/m^2, daunorubicin dose 118 mg/m^2, doxorubicin dose 181 mg/m^2, methotrexate dose 7934 mg/m^2, and VP 16 dose 6184 mg/m^2.

Table I.1 Information table for Example 1

Variable	Input	Variable	Input	Variable	Input
Race	White	BCNU	0	Carboplatin	0
Age at Cancer Diagnosis	2	Busulfan	0	Cis_Platinum	0
BMT Indicator	No	CCNU	0	Bleomycin	0
Cancer Diagnosis Type	Leukemia	Chlorambucil	0	Daunorubicin	49.3 (mg/m^2)
Minimum ovarian radiation dose	280 (cGy)	Cyclophosphamide	0	Doxorubicin	0
Radiation doses to pituitary	2410 (cGy)	Ifosfamide	0	Idarubicin	0
Total body irradiation dose	0	Melphalan	0	Methotrexate	9404 (mg/m^2)
		Nitrogen_Mustard	0	Mitoxantrone	0
		Procarbazine	0	VM 26	1298 (mg/m^2)
		Thiotepa	0	VP 16	0

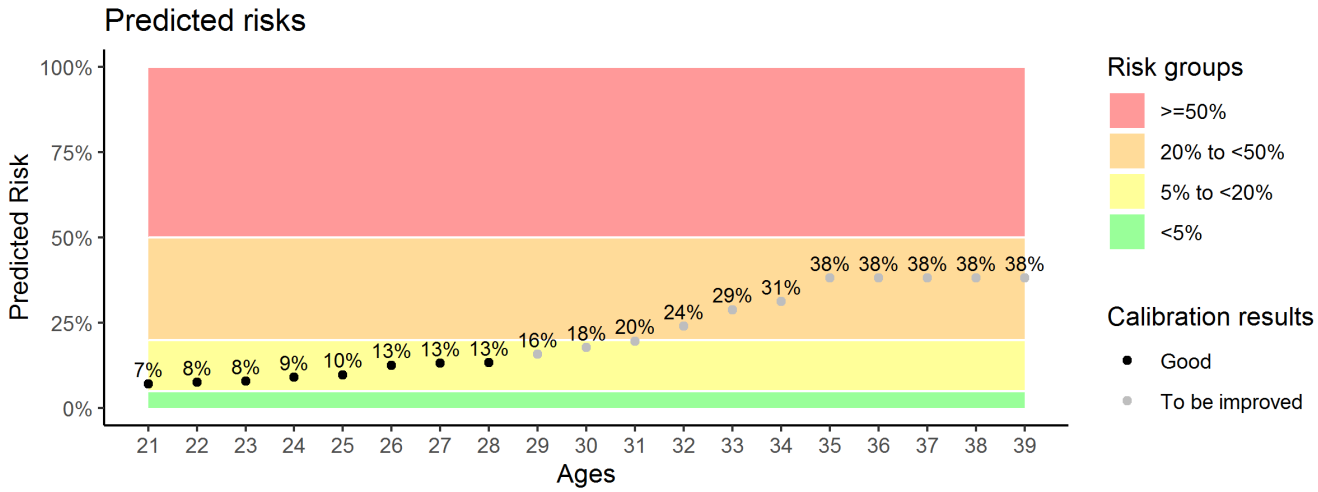


Figure I.1 Predicted risks of POI from age 21 to 39 for Example 1

Table I.2 Information table for Example 2

Variable	Input	Variable	Input	Variable	Input
Race	Asian or Pacific Islander	BCNU	300 (mg/m^2)	Carboplatin	0
Age at Cancer Diagnosis	14	Busulfan	0	Cis_Platinum	0
BMT Indicator	Yes	CCNU	0	Bleomycin	0
Cancer Diagnosis Type	NHL[a]	Chlorambucil	0	Daunorubicin	118 (mg/m^2)
Minimum ovarian radiation dose	2.3 (cGy)	Cyclophosphamide	15576 (mg/m^2)	Doxorubicin	181 (mg/m^2)
Radiation doses to pituitary	47 (cGy)	Ifosfamide	0	Idarubicin	0
Total body irradiation dose	0	Melphalan	0	Methotrexate	7934 (mg/m^2)
		Nitrogen_Mustard	0	Mitoxantrone	0
		Procarbazine	0	VM 26	0
		Thiotepa	0	VP 16	6184 (mg/m^2)

a: NHL: Non-Hodgkin lymphoma

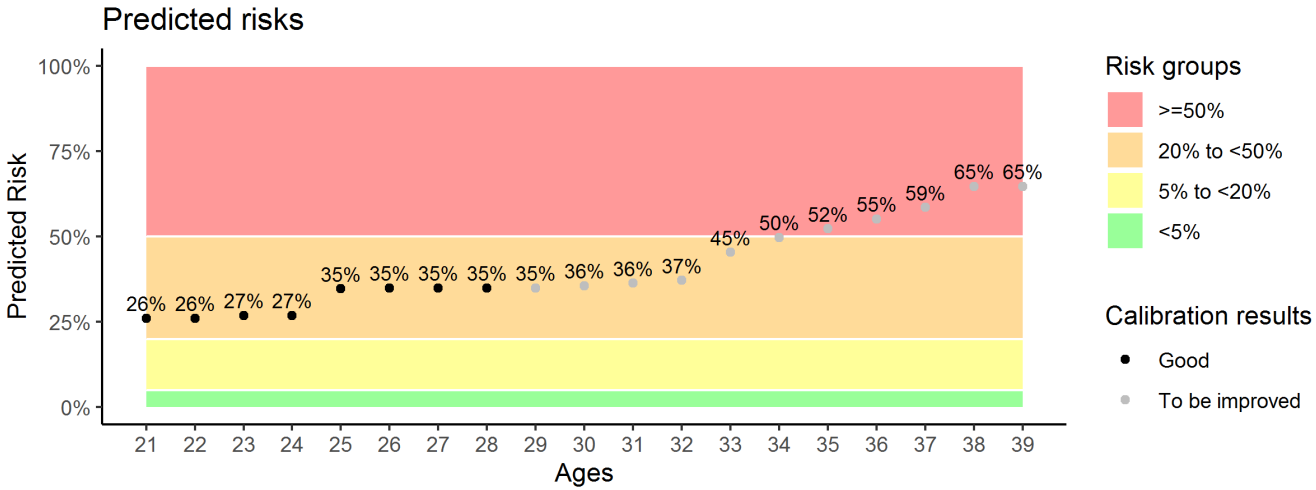


Figure I.2 Predicted risks of POI from age 21 to 39 for Example 2