Predicting Acute Ovarian Failure in Female Childhood Cancer Survivors:

A Cohort Study in the Childhood Cancer Survivor Study (CCSS) and

the St. Jude Lifetime Cohort (SJLIFE)

Authorship List Line

Rebecca A. Clark, Sogol Mostoufi-Moab, Yutaka Yasui, Ngoc Khanh Vu, Charles A. Sklar[†], Tarek Motan, Russell J. Brooke, Todd M. Gibson, Kevin C. Oeffinger[†], Rebecca M. Howell, Susan A. Smith, Zhe Lu, Leslie L. Robison, Wassim Chemaitilly, Melissa M. Hudson[†], Gregory T. Armstrong, Paul C. Nathan*[†], Yan Yuan*[✉].

[†] Full professor

* Co-senior authors

[✉] Corresponding author

List of Affiliations

School of Public Health - University of Alberta, Edmonton, Alberta, Canada. (RA Clark MSc, NK Vu PhD, Z Lu MEng, Y Yuan PhD)

University of Pennsylvania Perelman School of Medicine, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, United States of America. (S Mostoufi-Moab MD)

St. Jude Children's Research Hospital, Memphis, Tennessee, United States of America. (Y Yasui PhD, RJ Brooke PhD, TM Gibson PhD, LL Robison PhD, W Chemaitilly MD, MM Hudson MD, GT Armstrong MD)

Memorial Sloan Kettering Cancer Center, New York, New York, United States of America. (CA Sklar MD)

Faculty of Medicine and Dentistry - University of Alberta, Edmonton, Alberta, Canada. (T Motan MB ChB)

Duke University School of Medicine, Durham, North Carolina, United States of America. (KC Oeffinger MD)

The University of Texas MD Anderson Cancer Center, Houston, Texas, United States of America. (RM Howell PhD, SA Smith MPH)

The Hospital for Sick Children, Toronto, Ontario, Canada. (PC Nathan MD)

Corresponding Author Information

      Yan Yuan, PhD

      Associate Professor

      yyuan@ualberta.ca

      Tel: +1 (780) 248-5853

      School of Public Health, University of Alberta

      3-299 Edmonton Clinic Health Academy

      11405 87 Avenue

      Edmonton, Alberta

      T6G 1C9

      Canada

Word count: 4,123

Summary

Background

Cancer therapy can cause gonadal impairment. Acute ovarian failure (AOF) is defined as the permanent loss of ovarian function within five years of cancer diagnosis. We aimed to develop and validate risk prediction tools to provide accurate clinical guidance to paediatric cancer patients.

Methods

AOF risk prediction models were developed using eligible female participants in the Childhood Cancer Survivor Study (CCSS) cohort and validated in the St. Jude Lifetime Cohort (SJLIFE). Eligibility criteria were at least age 18, had complete treatment exposure and adequate menstrual history information available. Logistic regression, random forest, and support vector machines were used as candidate methods. Prediction performance was evaluated internally and externally using the areas under the ROC curve (AUC) and the precision-recall curve (AP). An online risk calculator was developed for clinical use.

Findings

Three-hundred and fifty-three (6%) of 5,886 CCSS participants and 50 (5.7%) of 875 SJLIFE participants experienced AOF. The median follow-up for the CCSS and SJLIFE analysis samples was 23·9 (IQR=20·4-27·9) and 23·9 (19·0-30·0) years, respectively. A prescribed dose model with abdominal and pelvic radiation doses and an ovarian dose model with ovarian radiation dosimetry using logistic regression were selected. Common predictors in both models were history of hematopoietic stem cell transplantation (HSCT), cumulative alkylating agent dose, and an interaction between age at cancer diagnosis and HSCT. External validation produced an estimated

AUC of 0·94 (95% CI=0·90-0·98) and AP of 0·68 (95% CI=0·53-0·81) for the ovarian dose model, and AUC of 0·96 (0·94-0·97) and AP of 0·46 (0·34-0·61) for the prescribed dose model.

Interpretations

Both AOF risk prediction models perform very well. The ovarian model is preferred if ovarian radiation dosimetry is available. The models, along with the online risk calculator, can aid clinical discussions regarding the need for fertility preservation interventions in young females newly diagnosed with cancer.

Funding

Research in Context

Evidence before this study

An increased risk of premature gonadal failure has been demonstrated in paediatric cancer survivors treated with chemotherapy and radiation. Six percent of female childhood cancer survivors lose ovarian function within five years of treatment (acute ovarian failure [AOF]), and an additional nine percent experience premature, non-surgical menopause before age 40. The time frame between primary cancer diagnosis and treatment resulting in AOF is limited to identify high-risk patients that will benefit from interventions aimed at fertility preservation. We searched PubMed with no date or language restrictions for all studies to evaluate the current knowledge of AOF and the associated risk factors in childhood cancer survivors using the search terms "pediatric cancer OR childhood cancer" AND "acute ovarian failure OR primary ovarian insufficiency" AND "risk". Five publications were considered for further review as they described AOF as an independent condition without grouping patients in a broader premature menopause umbrella. While high dose pelvic radiation, hematopoietic stem cell transplantation, and alkylating chemotherapy have been identified as risk factors associated with AOF, clinicians lack a tool that accurately estimates the risk of AOF for individual paediatric cancer patients at the time of cancer diagnosis. We did not find any study that aimed to develop risk estimates of AOF for individual paediatric cancer patients at the time of cancer diagnosis.

Added value of this study

To our knowledge, we have developed and validated the first models for predicting the risk of AOF in female childhood cancer survivors. While physicians are aware of the gonadotoxic treatment exposures with a high likelihood of causing AOF, there are no available prediction tools

to estimate the AOF risk for a given patient based on a planned oncologic treatment regimen. Having a precise risk estimate available to clinicians will guide informed discussions with patients and their families for time-sensitive interventions to preserve fertility function prior to initiation of cancer treatment and inform the need for future ovarian hormone replacement treatment after completion of cancer therapy. We provide an easily accessible and user-friendly, online AOF risk calculator for clinicians to directly calculate each patient's risk for AOF based on their planned cancer treatment. The developed models perform very well both internally and externally, highlighting the validity of the risk estimates and ensures clinical recommendations are provided with confidence.

Implications of all the available evidence

As the majority of childhood cancer patients will become long-term survivors, the focus of cancer survivorship research has shifted toward maximizing survivor quality of life. Our models and the associated web application can help inform discussions with female patients and their families at the time of cancer diagnosis regarding the need for fertility preservation prior to cancer treatment and the possible need for ovarian hormone replacement after completion of cancer therapy.

Introduction

Due to advancements in cancer treatment and supportive care, there are almost 500,000 survivors of childhood cancer in the United States,[1] and between 300,000 and 500,000 childhood cancer survivors in Europe.[2,3] Survivors are at an increased risk of developing chronic health conditions due to cancer treatment-related toxicities.[4] The cumulative burden of treatment-associated chronic health conditions is substantial, with survivors experiencing an average of 17 conditions by age 50 years.[5] Impaired gonadal function is a common late effect of cancer therapy that can have a significant impact on a survivor's quality of life.[6]

Females are born with a finite supply of ovarian follicles that decline with age.[7] Cancer-directed therapies such as radiation and alkylating chemotherapy can accelerate this decline, resulting in early cessation of ovarian endocrine and reproductive function.[7] Primary ovarian insufficiency (POI) is defined as compromised gonadal function prior to age 40.[8] POI can manifest as acute ovarian failure (AOF) or premature menopause (PM). AOF occurs when an individual permanently stops menstruating within five years of their cancer diagnosis, or fails to progress through puberty or to achieve menarche by 18 years of age following cancer treatment.[9] Prior investigations from the Childhood Cancer Survivor Study (CCSS) have demonstrated a 6% prevalence of AOF in female survivors of childhood cancer.[9] PM occurs when ovarian function is retained for at least five years following cancer diagnosis, but non-surgical menopause develops before age 40.[10] In the general population, the prevalence of premature, non-surgical menopause is approximately 1%,[11] whereas the cumulative incidence of PM (excluding AOF) reported in CCSS female survivors is 9% by age 40[10]. AOF and PM can restrict reproductive options, reduce quality of life, increase anxiety and depressive feelings, and increase the risk for serious morbidities including ischemic heart disease, osteoporosis, and cognitive decline.[10,12–15]

Better precision in prediction of an individual's risk for developing POI can facilitate appropriate counselling and fertility preservation at the time of cancer diagnosis and be used to evaluate the need for future hormone replacement therapies. Obtaining an accurate risk estimate is important as available fertility preservation technologies such as ovarian tissue and oocyte cryopreservation are expensive and invasive. Ovarian tissue cryopreservation requires surgery, the potential for future livebirths is poorly established, and there is concern that certain malignancies may involve the ovary, precluding reimplantation of the tissue. The operative risk may be elevated in children who are immunocompromised or have abnormal blood counts, increasing their risk for infection or bleeding. Oocyte harvest can only be offered to post pubertal females and requires time for ovarian hyperstimulation prior to oocyte retrieval impeding use before initiation of gonadotoxic treatment. No studies have addressed the safety or success rate of these procedures in children and adolescents.[16,17]

While physicians are aware of the gonadotoxic treatment exposures likely to cause AOF, there are no available prediction tools to estimate the AOF risk for a given patient based on a planned oncologic treatment regimen. To address this gap, we developed and externally validated risk prediction models for clinical use. Our goal was to focus on risk prediction of AOF at the time of cancer diagnosis to inform clinicians of the need for time-sensitive interventions. By offering fertility preservation procedures prior to treatment initiation to patients at a significant risk of AOF, the opportunity for future reproduction in this group can be maximised and the risk of performing unnecessary procedures for low AOF risk patients can be minimised.

Methods

Study design and participants

The CCSS, a multi-institutional longitudinal cohort study of 24,362 survivors of childhood cancer from North America, was the primary source of data. Originally established in 1994, its cohort characteristics, eligibility criteria, and study design features have been documented elsewhere.[18,19] Briefly, the cohort includes five-year survivors diagnosed before age 21 with an eligible cancer diagnosis (leukaemia, central nervous system cancers, Hodgkin lymphoma, non-Hodgkin lymphoma, Wilms tumour, neuroblastoma, soft tissue sarcoma, or bone tumours) treated at one of 31 participating North American institutions between January 1, 1970 and December 31, 1999. Survivors were eligible for the AOF prediction analysis if they were female, had complete treatment exposure data, were at least 18 years of age at their latest follow-up, and provided menstrual history information including age at menarche, current menstrual status, age at last menstrual period, and etiology of menopause (surgical vs. nonsurgical) for those who were currently menopausal.[9]

The validation cohort was derived from the St. Jude Lifetime Cohort (SJLIFE) which was established in September 2007 to medically assess the health status of childhood cancer survivors treated at St. Jude Children's Research Hospital.[4,20] Participants were eligible for SJLIFE if they had been diagnosed and treated for a paediatric cancer at St. Jude Children's Research Hospital after 1962 and were at least ten-year survivors.[4,20] Participants were excluded from the AOF prediction analysis if they had an ovarian hormone deficiency, were missing treatment exposure information, or if their ovarian status was unable to be determined.[8] Individuals who participated in both cohorts (CCSS and SJLIFE) were excluded from model development using CCSS and used only in the SJLIFE validation analysis.

Ethics approval for the study was obtained from the University of Alberta Health Research Ethics Board (PRO00067066). The institutional review board for each participating CCSS institutions approved the CCSS protocol, the St. Jude Children's Research Hospital approved the SJLIFE protocol, and informed consent was obtained from all study participants.

Procedures

Self-reported demographic and outcome information was obtained from CCSS participants through baseline and follow-up questionnaires, and treatment exposure information was abstracted from medical and radiation records.[19] For this analysis, radiation doses, including all therapy received within 5 years of primary cancer diagnosis, were estimated to the ovaries and two body regions (abdomen and pelvis) with the potential to overlap with the ovaries.[21] To obtain the ovarian radiation dose, the average doses to the right and left ovaries were estimated separately, and the minimum of the two values was used in the analysis. For the body regions, the maximum target dose was determined by summing the prescribed dose from all overlapping fields in each of the respective regions. Ovarian radiation and alkylating agent doses used during validation were calculated as in Chemaitilly et al, 2017.[8]

Classification of ovarian status, the primary outcome, was assigned to CCSS participants using an established definition,[6] or manually by endocrinologists (SM-M, CAS) for ambiguous cases, based on responses to menstrual history questions on baseline and/or follow-up questionnaires. Individuals who achieved menarche prior to cancer treatment were classified with AOF if they failed to resume menstruating within five years of treatment. Individuals who had not experienced menarche were classified with AOF if menarche was not achieved by age 18.[9] Age 18 was selected to avoid misclassifying patients as AOF that may not have achieved menarche by age 15 or 16

years due to cancer-related issues such as poor nutrition and weight loss but went on to achieve menarche a year or two later. The remaining individuals in the study sample were classified as not having developed AOF. As AOF is an acute event developed during or shortly after treatment completion, the AOF status was assessable at the survey completed after survivors turned 18 and analysis methods for a binary outcome were used. Among survivors who reported taking an oral contraceptive pill (OCP), the survey differentiated between those still menstruating but taking an OCP for contraception, those taking an OCP or hormone supplements to regulate their menstrual cycles, and those taking an OCP or hormone supplements as ovarian replacement therapy. Ovarian status classifications for SJLIFE participants were ascertained by an endocrinologist (WC) based on questionnaire responses and hormone measurements.[22,23] This allowed for evaluation of the prediction algorithms on clinically verified ovarian status classifications in this independent validation sample.

Potential predictors for model development included age at cancer diagnosis, age at menarche, cancer diagnosis, any exposure to chemotherapy, cumulative dose of alkylating agents measured using the cyclophosphamide equivalent dose (CED),[24] hematopoietic stem cell transplant (HSCT), and ovarian radiation doses for an ovarian dose model. As ovarian radiation dose information is not universally available, we also considered a prescribed dose model where the ovarian radiation dose term was replaced with protocol-specified abdominal and pelvic radiation doses while retaining all other variables. Prediction algorithms used for model development included logistic regression, random forest,[25] and support vector machines.[26] Clinician input and model performance were utilised to select the best model for each method.

Model performance was evaluated both internally and externally. To avoid over-optimistic estimates of performance during internal validation, the eligible CCSS data was divided into

"training" (75%) and "test" (25%) sets. A model was built on the training set using the variables selected in the model development stage, and the AOF risk was predicted using this model for subjects in the corresponding test set. We repeated this process 100 times on 100 randomly split training and test sets and took the average of the predicted risk for each subject. Model performance was evaluated using the average predicted risk and the observed AOF status.

Along with the continuous risk estimate, we wanted to provide users with an optional categorisation of the AOF risk from the model. Predicted risks from the best models were used to stratify patients into low (< 5%), medium-low (5-20%), medium (20-50%), and high risk ($\geq$ 50%) categories. Risk value thresholds for the categories were defined by two endocrinologists (SM-M, CAS) and a paediatric oncologist (PCN) with expertise in childhood cancer survivorship such that the groups represented reasonable and clinically meaningful categories for oncofertility discussions and decision making.

Outcomes

The primary outcome in our analysis was acute ovarian failure, which is defined as the permanent loss of ovarian function within five years of cancer diagnosis or failure to achieve menarche by age 18 years following cancer treatment.

Statistical analysis

The ovarian status was deemed not assessable for participants who were exposed to a pituitary radiation dose > 30 Gray (Gy), who had a tumour in the hypothalamus/pituitary region, or whose menstrual history information was incomplete, unclear, or provided by a proxy. Thus, these survivors were excluded from the analysis. Survivors who had missing exposure of radiation

and/or chemotherapy were also excluded. A brief overview of the statistical methodology is provided; a detailed description is available in the Supplementary Material. We developed candidate prediction algorithms using three popular methods for binary outcome analysis (logistic regression, random forest, and support vector machines), which allows for examination of the sensitivity of the data to the modelling method. The best ovarian dose and prescribed dose models were selected from these candidate prediction algorithms and subsequently externally validated using the SJLIFE data. Analysis was performed using Stata version 14·2, R version 3·4·3, and SAS version 14·1.

Receiver operating characteristic (ROC) curves and area under the ROC curve (AUC) values were used to evaluate the ability of the model to distinguish AOF cases from non-AOF cases when presented with a pair of observations, one with AOF and the other without. AUC values range from 0·5 to 1·0, with values closer to 1·0 being preferred. Precision-recall (PR) curves and area under the PR curve, a.k.a. average positive predictive values (AP) were used to measure the ability of the model to detect AOF cases from the entire population of female childhood cancer patients.[27–29] The positive predictive value represents the probability that a patient classified as AOF is truly an AOF patient. The AP value can be interpreted as the AOF risk for a patient whose predicted risk is greater than the risk estimate of a randomly selected female survivor with AOF. AP values range from the value of the event rate in the population to 1·0, and when comparing candidate models on specific populations, larger values indicate superior performance. Scaled Brier scores were reported for overall model performance evaluation (larger values are preferred when comparing models)[30] and calibration curves were used to visually inspect the alignment of the predicted AOF risk with the observed AOF risk.[30]

Role of the Funding Source

The funding source did not have any role in the study design, collection, analysis, or interpretation of the data, in the writing of the report, or in the decision to submit the paper for publication. The corresponding author had full access to the CCSS analysis cohort data used in the study and had final responsibility for the decision to submit for publication.

Results

Of the 11,336 total female CCSS participants, 5,450 (48·1%) were excluded. A detailed flowchart of the CCSS sample and reasons for exclusion is shown in Figure 1. From the 1,644 total female survivors in the SJLIFE cohort, 723 (44·0%) were excluded as in Chemaitilly et al, 2017,[8] and an additional 46 survivors were excluded due to missing radiation data (n = 38) or a subsequent malignancy (n = 8). Data was collected for participants in the CCSS sample between November 3, 1992 and November 25, 2016, and between October 17, 2007 and April 16, 2012 for the SJLIFE sample.

Demographic, diagnostic, and treatment characteristics of the 5,886 eligible CCSS survivors and the 875 eligible SJLIFE survivors are presented in Table 1. Almost one-third (1,869 (31·8%) of 5,886) of the CCSS survivors had been diagnosed with leukaemia, and 3·7% (217 of 5,886) underwent a HSCT. Survivors in the CCSS analysis cohort had an overall median of 23·9 (IQR = 20·4-27·9) years of follow-up and a total follow-up of 142,738·8 person-years. For survivors in the SJLIFE analysis cohort, the median was 23·9 (19·0-30·0) years and the total follow-up was 21,492·2 person-years. As AOF is an acute event, every survivor's AOF status in these two cohorts was completely ascertained at either 18 years of age or 5 years post cancer diagnosis, depending

on her menarcheal status at cancer diagnosis. Thus, the effective follow-up time for determining

AOF was a median of 5·0 (IQR 5·0-9·1) years and a total follow-up of 42,335·1 person-years in

the CCSS analysis cohort and 5·0 (IQR 5·0-8·8) and 6,150·7 person-years in the SJLIFE analysis

cohort. Three hundred and fifty-three of 5,886 survivors were classified with AOF in the CCSS

sample, corresponding to a prevalence of 6·0%. A similar prevalence was observed for the SJLIFE

survivors, where 50 survivors of 875 (5·7%) were diagnosed with AOF. Among survivors in the

CCSS cohort, 84 had initiated OCP use within 5 years of their cancer diagnosis and were classified

as not having AOF after detailed review of their survey responses by two endocrinologists (SM-

M, CAS).

Best models from the three candidate methods (logistic regression, random forest, and support

vector machines) performed similarly (Supplementary Material page 6). Therefore, the logistic

regression models were selected due to the transparency and interpretability of logistic regression

compared to the other methods. Variables and their estimated coefficients for the best ovarian dose

model and prescribed dose model are shown on pages 4-5 of the Supplementary Material.

ROC curves of the best risk prediction models are shown in Figure 2a and Figure 2b. When the

CCSS test sets were used for internal evaluation, the ovarian dose model AUC value was 0·82

(95% CI = 0·79-0·85) and the prescribed dose model AUC value was 0·78 (95% CI = 0·74-0·81).

The AUC values for both models increased when the SJLIFE cohort was used as an independent

validation set. The ovarian dose model AUC value was 0·94 (95% CI = 0·90-0·98), and the

prescribed dose model AUC value was 0·96 (95% CI = 0·94-0·97).

PR curves of the best models are presented in Figure 2c and Figure 2d. For internal validation

using the CCSS test sets, the AP value was 0·50 (95% CI = 0·45-0·56) for the ovarian dose model

and 0·37 (95% CI = 0·32-0·43) for the prescribed dose model. Similar to the AUC results, the AP

values increased when externally validated in SJLIFE. The AP was 0·68 (95% CI = 0·53-0·81) for the ovarian dose model and 0·46 (95% CI = 0·34-0·61) for the prescribed dose model.

The scaled Brier score for the ovarian dose model was 31·4% in the CCSS cohort and 49·9% in the SJLIFE validation cohort. The scaled Brier score for the prescribed dose model remained relatively consistent at 20·0% for the CCSS cohort and 22·6% for the SJLIFE cohort.

The calibration curves for the CCSS cohort (Figure 3a) closely follow the diagonal for both the ovarian dose and prescribed dose model, indicating good alignment between the observed and predicted risk. For the SJLIFE cohort, while the majority of the calibration curves (Figure 3b) follow the diagonal line, a slight deviation toward the end of the curves suggests that the models are not as well calibrated at the lower end of the risk, indicating less reliable risk estimates.

Table 2 and Table 3 present the results of categorising CCSS and SJLIFE patients into the predefined risk categories for the ovarian dose model and prescribed dose model, respectively. The ovarian dose model categorised survivors more precisely and accurately than the prescribed dose model, as seen by the larger counts in the high (≥ 50%) and low (< 5%) risk groups. Specifically, using the ovarian dose model, 5,130 (87·2%) of 5,886 participants in the CCSS cohort were estimated to be at low risk, whereas the prescribed dose model classified 4,898 (83·2%) of 5,886 individuals as low risk. On the other end of the predicted risk spectrum, the ovarian dose model classified 182 individuals as high risk (of whom 132 were AOF cases, 72·5%), while the prescribed dose model predicted only 97 individuals as high risk with 61 (62·9%) having developed AOF. A cross-table of the risk estimates from the prescribed dose and ovarian dose models based on CCSS cohort and a detailed comparison is included in the Supplementary Material (pages 2, 5).

Discussion

With AUC values from 0·78 to 0·82 in the CCSS cohort and 0·94 to 0·96 in the external validation using the SJLIFE cohort, we have developed, to our knowledge, the first risk prediction models for AOF that provide a high level of confidence appropriate for use in a clinical setting. As AP values larger than the population event rate imply superior predictive ability to detect AOF cases, values for the prescribed dose and ovarian dose models that range between 0·37 and 0·68 indicate outstanding predictive power for detecting AOF compared to the event rate of 0·06 in both cohorts. Given the limited time interval between cancer diagnosis, treatment, and the subsequent development of AOF, it is crucial to appropriately counsel high-risk patients and reassure low-risk patients at the time of cancer diagnosis. Our goal was to develop and validate an easily accessible and user-friendly clinical tool to aid clinicians at the time of cancer diagnosis by providing personalised risk assessments of future ovarian function for patients.

The outstanding performance in the external SJLIFE cohort further confirms that our prediction algorithms are generalisable. The main outcome for our risk prediction model was AOF, which was classified using self-reported menstrual history information provided in CCSS questionnaires. CCSS participant ovarian status was not verified clinically, and thus subject to potential misclassification. The SJLIFE cohort has clinically verified ovarian status classifications, with menstrual history information provided by the participant supplemented by ovarian hormone levels, permitting more precise ascertainment of ovarian status.[8,22,23] Although we developed the models using the outcome data with a higher potential for misclassification, we observed an increase in the predictive ability in the SJLIFE cohort, which highlights the robustness of our models. Survivors in the SJLIFE cohort were exposed, on average, to higher alkylating agent doses than survivors in CCSS, but fewer received radiation to the ovaries. This likely reflects differences

17

in the distribution of diagnoses between the two cohorts, in addition to the fact that some patients in SJLIFE were treated in a more recent era where, in an effort to improve late health outcomes, there have been attempts to reduce radiation exposure by increasing the use of certain chemotherapy agents. Despite these differences, the risk models perform well in both cohorts, lending support to their generalisability.

When the predicted risks from the ovarian dose and prescribed dose models were stratified into risk categories, the ovarian dose model was superior to the prescribed dose model. The ovarian dose model classified more subjects into the low and high-risk categories, with a smaller number of individuals categorised into the intermediate risk categories; categories which have greater uncertainty regarding the necessity of intervention. However, in order to use the ovarian dose model, the estimated radiation dose to the ovaries is required. This information may not be available to all providers. In contrast, the radiation doses used in the prescribed dose model can be derived from the planned abdominal and pelvic target doses without the need for sophisticated dosimetry calculations. Performance evaluation of the prescribed dose model confirms that accurate and reliable risk estimates are still obtained, and clinicians can feel confident using either model for AOF risk prediction.

A web-based application of the AOF models is under development for clinical use and will be available at ccss.stjude.org/aofcalc. To utilise this application, clinicians input the patient's age at cancer diagnosis and proposed treatment exposures (whether they will undergo a HSCT, the total body irradiation dose, any additional abdominal and pelvic or ovarian radiation doses, as well as either the CED value or specific alkylating agent doses). A continuous AOF risk estimate and corresponding risk category is calculated from the ovarian dose and/or prescribed dose model depending on the radiation information provided.

Validation of the risk models in a prospective cohort would further establish the validity of our tool. This is not currently planned, but since the risk prediction tool will be freely available online, we would encourage other investigators to apply it prospectively to their patient cohorts and publish their results. Importantly, the tool and its estimate of risk should not be the only criterion for discussing AOF risk with patients and families. Since some patients considered at low or moderate risk for AOF at first presentation of cancer might nonetheless develop AOF, while others will relapse and require therapy that increases their risk, all newly diagnosed patients with cancer should be counselled about the options for fertility preservation. Further, it is important that cancer survivors deemed to be at high risk for AOF do not assume that they will develop ovarian failure and use appropriate contraception to prevent unplanned pregnancies and sexually transmitted infections.

Although the developed models perform very well at predicting AOF cases in the population of childhood cancer survivors, there remain limitations. Cases of AOF may be masked in female survivors who are taking oral contraceptives or other hormone medications that result in persistent menstruation and may result in an underestimate of the prevalence of AOF. Eighty-four survivors in the CCSS cohort who were classified as non-AOF started OCP use within 5 years of their cancer diagnosis and were thus at risk for misclassification, but detailed review by the two study endocrinologists minimized this risk. The reliance on self-reported menstrual history and possible misclassification of patients at risk for both primary and central hypogonadism represent additional limitations; the fact that the models perform well in the SJLIFE cohort, which uses clinically verified hypogonadism and ovarian status in amenorrhoeic survivors less than 40 years of age, provides reassurance.

The models only predict the risk for AOF, and not for any other POI (such as PM), which may occur before age 40.[10] Predicting PM using statistical techniques is inherently more challenging since in addition to determining if it will occur, there is the added aspect of when it will occur. PM develops later than AOF, and variation in genetic susceptibility to the gonadotoxic treatments makes it difficult to generalise the PM risk.[23] Despite some of these challenges, we are currently developing models for PM and aim to produce a similar risk prediction tool for clinical use that generates age-specific PM risk.

A complete case analysis approach was used rather than multiple imputation in our study, which can result in bias and inefficiency. However, the excellent prediction performance of our models on an external cohort, and the models' good precision in the estimates of accuracy (AUC and AP), suggest that the complete case approach did not introduce meaningful bias, and our models are generalisable. Finally, none of the patients in our cohort were exposed to newer cancer treatments such as immunotherapy or targeted therapies. The risk prediction models will require updating as information about the impact of such therapies on ovarian function emerges.

As the majority of childhood cancer patients will become long-term survivors, the focus of cancer survivorship research has shifted toward maximizing survivor quality of life. Our models and the associated web application can help inform discussions with female patients and their families at the time of cancer diagnosis regarding the need for fertility preservation prior to cancer treatment and the possible need for ovarian hormone replacement after completion of cancer therapy.

Abbreviations

AOF   Acute ovarian failure

CED   Cyclophosphamide equivalent dose

CCSS   Childhood Cancer Survivor Study

HSCT   Hematopoietic stem cell transplant

OCP   Oral contraceptive pill

PM   Premature menopause

POI   Primary ovarian insufficiency

SJLIFE   St. Jude Lifetime Cohort

Declaration of Interests

Other Locations Where This Study Has Been Presented

Presented as an oral presentation at the North American Symposium on Late Complications after Childhood Cancer (NASLCCC) in Atlanta, GA (June 22, 2019).

Author Contributions

Conception: YYuan, PC, KCO, YYasui, SM-M, CAS, TM

Data acquisition: YYasui, RAC, GTA, SM-M, CAS, TMG, RMH, SAS, LL, WC, MMH, YYuan

Study design: YYasui, GTA, KCO, TMG, LLR, MMH, YYuan

Data preparation and analysis: RAC, NKV, RJB, ZL, YYuan, YYasui

Figures: NKV, ZL, RJB, YYuan

Interpretation of results: RAC, NKV, ZL, YYuan, YYasui, PCN, SM-M, CAS, WC

Manuscript drafting and revision: RAC, SM-M, PCN, YYuan, YYasui, CAS, RMH, SAS, RJB, MMH, GTA, WC, LLR, TM, NKV, TMG, KCO, ZL

Acknowledgements

Table 1: Diagnostic and treatment characteristics of the CCSS study cohort and the SJLIFE validation cohort

	CCSS (n = 5,886) n (%)	SJLIFE (n = 875) n (%)
Age at Cancer Diagnosis, years of age		
Median (25% - 75% percentiles)	7·28 (3·21 - 13·71)	6·70 (3·38 - 13·21)
Cancer Diagnosis		
Leukaemia	1,869 (31·8)	378 (43·2)
Hodgkin lymphoma	817 (13·9)	122 (13·9)
Kidney tumours	755 (12·8)	67 (7·7)
Bone cancer	630 (10·7)	56 (6·4)
Central nervous system	616 (10·5)	44 (5·0)
Neuroblastoma	489 (8·3)	42 (4·8)
Non-Hodgkin lymphoma	379 (6·4)	38 (4·3)
Soft tissue sarcoma	331 (5·6)	47 (5·4)
Other*	NA	81 (9·3)
Cumulative Alkylating Agent Dose (CED Value, g/m^2)		
None	3,130 (53·2)	363 (41·5)
< 4	801 (13·6)	75 (8·6)
4 to < 8	743 (12·6)	163 (18·6)
≥ 8	1,212 (20·6)	274 (31·3)
Ovarian Radiation Dose, Gy		
None	3,174 (53·9)	710 (81·1)
< 10	2,383 (40·5)	108 (12·3)
10 to < 20	210 (3·6)	30 (3·4)
≥ 20	119 (2·0)	27 (3·1)
Abdominal Radiation Dose, Gy		
None	3,173 (53·9)	368 (42·1)
< 10	1581 (26·9)	325 (37·1)
10 to < 20	379 (6·4)	55 (6·3)
≥ 20	753 (12·8)	127 (14·5)
Pelvic Radiation Dose, Gy		
None	3,173 (53·9)	368 (42·1)
< 10	1883 (32)	345 (39·4)
10 to < 20	300 (5·1)	50 (5·7)
≥ 20	530 (9·0)	112 (12·8)
Hematopoietic Stem Cell Transplant		
Yes	217 (3·7)	18 (2·1)
No	5,669 (96·3)	857 (97·9)

*The SJLIFE cohort has less restrictive inclusion criteria, and as such, some patients did not fit into the restricted diagnostic categories for CCSS patients. We grouped these individuals together in the "other" cancer diagnosis category and retained them in the SJLIFE validation cohort to assess the algorithms' performance in a broader population.

Abbreviations: CCSS is the Childhood Cancer Survivor Study, SJLIFE is the St. Jude Lifetime Cohort, CED is the cyclophosphamide equivalence dose, and Gy is Gray.

**Table 2: Acute ovarian failure risk categories and cases as predicted by the best ovarian dose model**

Risk Category	Cohort	Survivors (n)	AOF Cases (n)	AOF %
Low (< 5%)	CCSS	5,130	119	2·3
	SJLIFE	796	8	1·0
Medium Low (5% - < 20%)	CCSS	429	47	11·0
	SJLIFE	34	8	23·5
Medium (20% - < 50%)	CCSS	145	55	37·9
	SJLIFE	8	4	50·0
High (≥ 50%)	CCSS	182	132	72·5
	SJLIFE	37	30	81·1

Abbreviations: AOF is acute ovarian failure, CCSS is the Childhood Cancer Survivor Study, and SJLIFE is the St. Jude Lifetime Cohort.

Table 3: **Acute ovarian failure risk categories and cases as predicted by the best prescribed dose model**

| Risk Category	Cohort	Survivors (n)	AOF Cases (n)	AOF %
Low (< 5%)	CCSS	4,898	117	2·4
	SJLIFE	709	1	0·1
Medium Low (5% - < 20%)	CCSS	515	62	12·0
	SJLIFE	114	26	22·8
Medium (20% - < 50%)	CCSS	376	113	30·1
	SJLIFE	39	17	43·6
High (≥ 50%)	CCSS	97	61	62·9
	SJLIFE	13	6	46·2

Abbreviations: AOF is acute ovarian failure, CCSS is the Childhood Cancer Survivor Study, and SJLIFE is the St. Jude Lifetime Cohort.
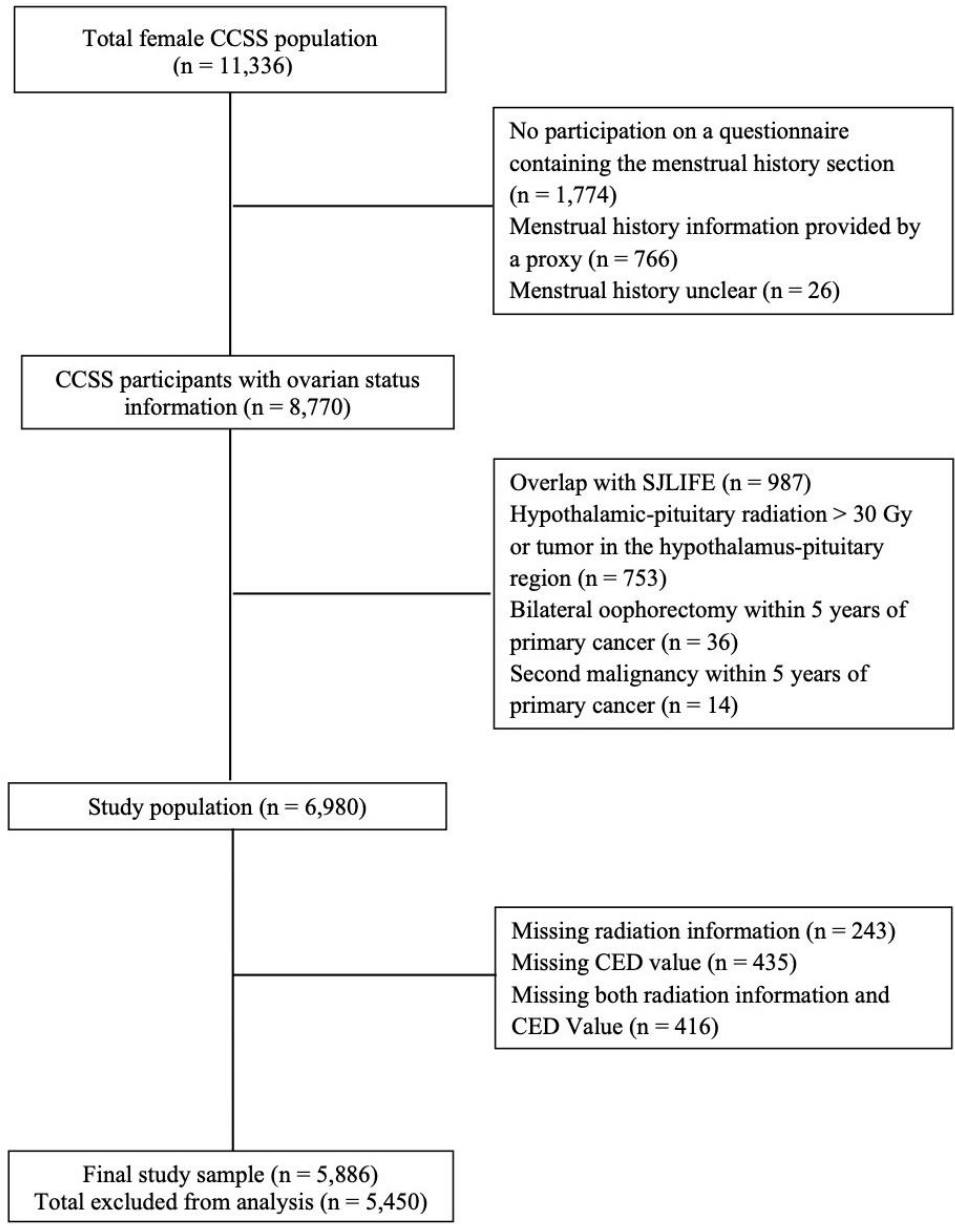
Figure 1: CCSS study sample flowchart for inclusion in model development

The reason for non-participation is unknown.

Abbreviations: CCSS is the Childhood Cancer Survivor Study, SJLIFE is the St Jude Lifetime Cohort Study, CED is the cyclophosphamide equivalent dose.
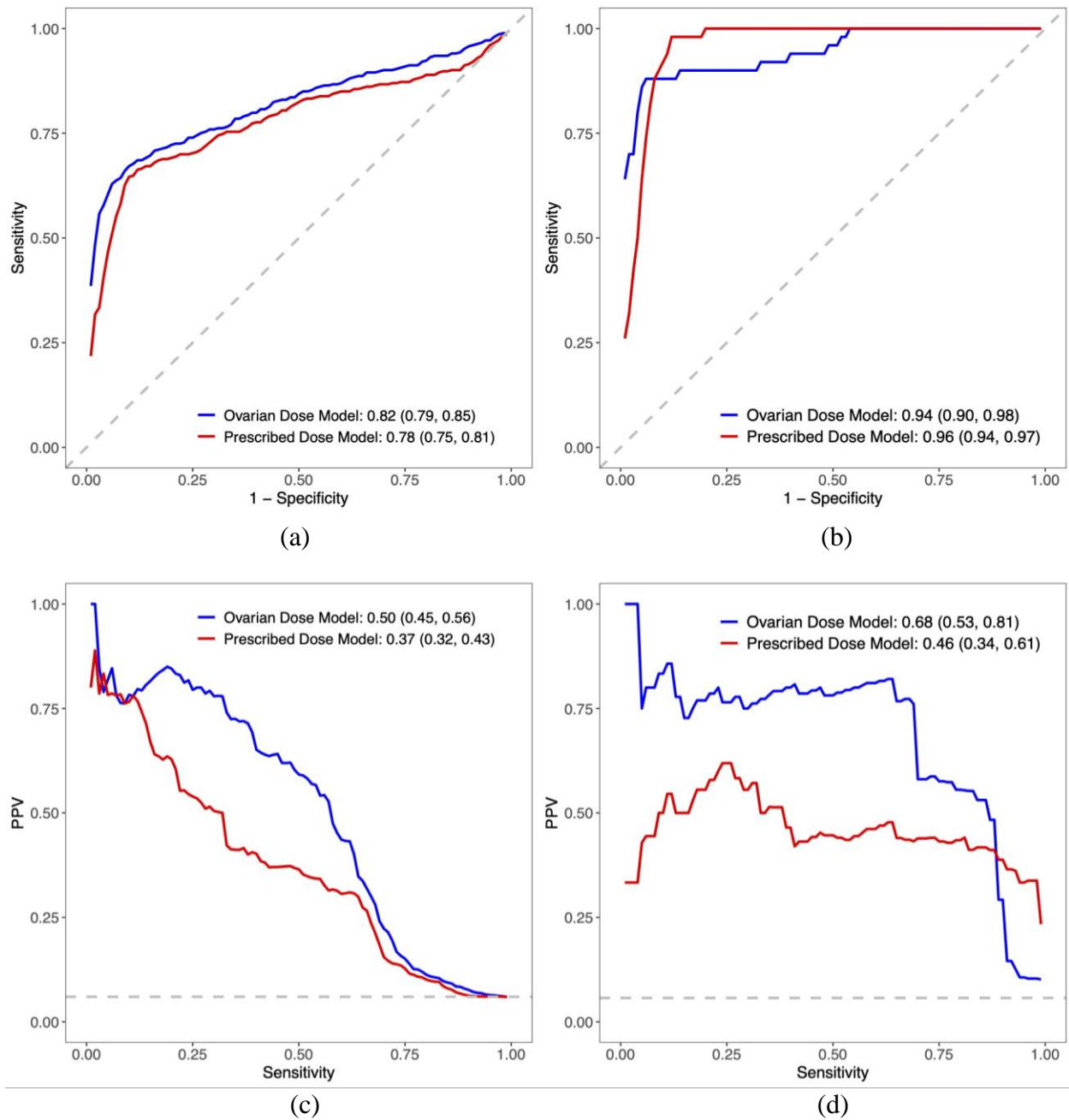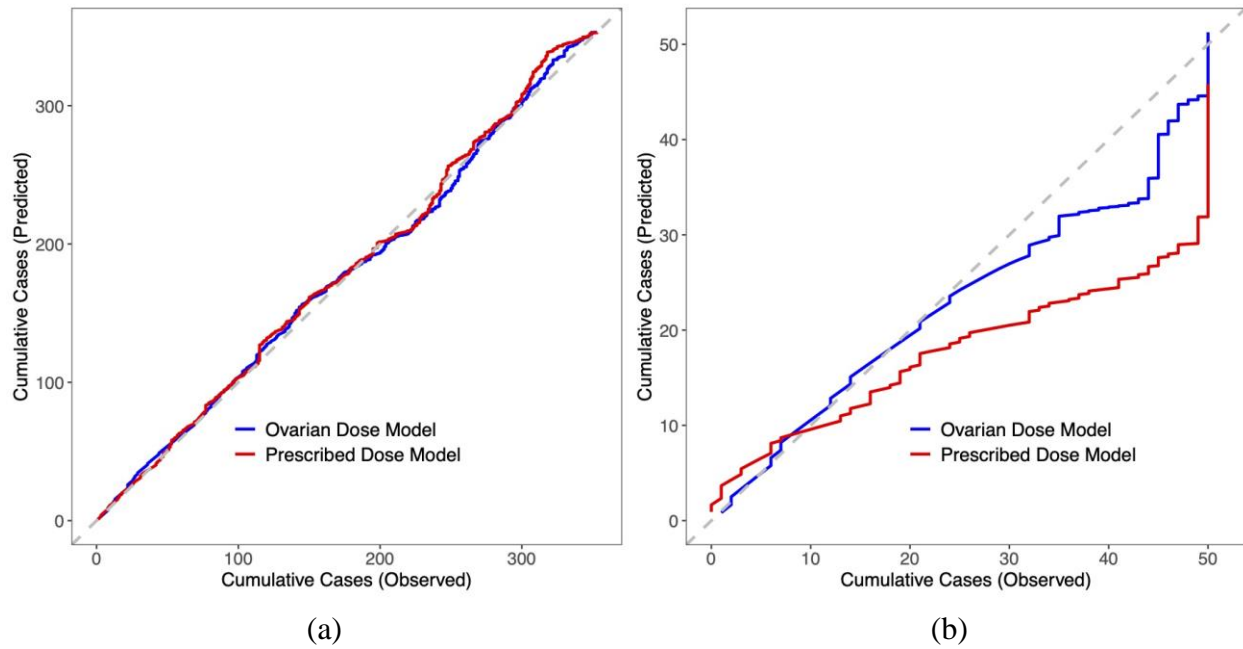
Figure 2: ROC and PR curves for the best models

The top row of the figure presents ROC curves and AUC values (95% CI) for the ovarian dose

model (blue) and the prescribed dose model (red), from the CCSS cohort (a) and the SJLIFE cohort

(b). The bottom row of the figure includes PR curves and AP values (95% CI) for the ovarian dose

model (blue) and the prescribed dose model (red), from the CCSS cohort (c) and the SJLIFE cohort (d).

Abbreviations: PPV is the positive predictive value, ROC is receiver operating characteristic curve, AUC is the area under the ROC curve, CI is the confidence interval, PR is the precision-recall, AP is the average positive predictive value, CCSS is the Childhood Cancer Survivor Study, and SJLIFE is the St Jude Lifetime Cohort.

(a)	(b)

Figure 3: Calibration curves for the best models

Calibration curves for the ovarian dose model (blue) and the prescribed dose model (red) in the

CCSS cohort (a) and the SJLIFE cohort (b).

Abbreviations: CCSS is the Childhood Cancer Survivor Study and SJLIFE is the St. Jude Lifetime

Cohort.

References

1 Robison LL, Hudson MM. Survivors of childhood and adolescent cancer: life-long risks and responsibilities. *Nat Rev Cancer* 2014; **14**: 61–70.

2 Hjorth L, Haupt R, Skinner R, *et al.* Survivorship after childhood cancer: PanCare: a European Network to promote optimal long-term care. *Eur J Cancer* 2015; **51**: 1203–11.

3 Haupt R, Essiaf S, Dellacasa C, *et al.* The 'Survivorship Passport' for childhood cancer survivors. *Eur J Cancer* 2018; **102**: 69–81.

4 Hudson MM, Ehrhardt MJ, Bhakta N, *et al.* Approach for Classification and Severity Grading of Long-term and Late-Onset Health Events among Childhood Cancer Survivors in the St. Jude Lifetime Cohort. *Cancer Epidemiol Biomarkers Prev* 2017; **26**: 666–74.

5 Bhakta N, Liu Q, Ness KK, *et al.* The cumulative burden of surviving childhood cancer: an initial report from the St Jude Lifetime Cohort Study (SJLIFE). *Lancet* 2017; **390**: 2569–82.

6 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**: 3240–7.

7 Johnston RJ, Wallace WHB. Normal ovarian function and assessment of ovarian reserve in the survivor of childhood cancer. *Pediatr Blood Cancer* 2009; **53**: 296–302.

8 Chemaitilly W, Li Z, Krasin MJ, *et al.* Premature Ovarian Insufficiency in Childhood Cancer Survivors: A Report from the St. Jude Lifetime Cohort. *J Clin Endocrinol Metab* 2017; **102**: 2242–50.

9 Chemaitilly W, Mertens AC, Mitby P, *et al.* Acute Ovarian Failure in the Childhood Cancer Survivor Study. *J Clin Endocrinol Metab* 2006; **91**: 1723–8.

10 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. *Cancer* 2018; **124**: 1044–52.

11   Torrealday S, Pal L. Premature Menopause. *Endocrinol Metab Clin North Am* 2015; **44**: 543–57.

12   Cousineau TM, Domar AD. Psychological impact of infertility. *Best Pract Res Obstet Gynaecol* 2007; **21**: 293–308.

13   Singer D, Mann E, Hunter MS, Pitkin J, Panay N. The silent grief: psychosocial aspects of premature ovarian failure. *Climacteric* 2011; **14**: 428–37.

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

15   Ryan J, Scali J, Carrière I, *et al.* Impact of a premature menopause on cognitive function in later life. *BJOG An Int J Obstet Gynaecol* 2014; **121**: 1729–39.

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

17   Corkum KS, Rhee DS, Wafford QE, *et al.* Fertility and hormone preservation and restoration for female children and adolescents receiving gonadotoxic cancer treatments: A systematic review. *J Pediatr Surg* 2019. DOI:10.1016/j.jpedsurg.2018.12.021.

18   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**: 2308–18.

19   Robison LL, Mertens AC, Boice JD, *et al.* Study design and cohort characteristics of the

Childhood Cancer Survivor Study: A multi- institutional collaborative project. *Med Pediatr Oncol* 2002; **38**: 229–39.

20   St. Jude Children's Research Hospital. Establishment of a Lifetime of Cohort of Adults Surviving Childhood Cancer (SJLIFE). 2018; **2019**. https://clinicaltrials.gov/ct2/show/study/NCT00760656.

21   Howell RM, Smith SA, Weathers RE, Kry SF, Stovall M. Adaptations to a Generalized Radiation Dose Reconstruction Methodology for Use in Epidemiologic Studies: An Update from the MD Anderson Late Effect Group. *Radiat Res* 2019; **192**: 169–88.

22   Hudson MM, Ness KK, Gurney JG, *et al.* Clinical Ascertainment of Health Outcomes Among Adults Treated for Childhood Cancer. *JAMA* 2013; **309**: 2371–81.

23   Brooke RJ, Im C, Wilson CL, *et al.* A High-Risk Haplotype for Premature Menopause in Childhood Cancer Survivors Exposed to Gonadotoxic Therapy. *J Natl Cancer Inst* 2018; **110**: 895–904.

24   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. *Pediatr Blood Cancer* 2014; **61**: 53–67.

25   Breiman L. Random Forests. *Mach Learn* 2001; **45**: 5–32.

26   Karatzoglou A, Meyer D, Hornik K. Support Vector Machines in R. *J Stat Softw* 2006; **15**: 1–28.

27   Ozenne B, Subtil F, Maucort-Boulch D. The precision–recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases. *J Clin Epidemiol* 2015; **68**: 855–9.

28   Yuan Y, Su W, Zhu M. Threshold-Free Measures for Assessing the Performance of

Medical Screening Tests. *Front public Heal* 2015; **3**: 57.

29   Yuan Y, Zhou QM, Li B, Cai H, Chow EJ, Armstrong GT. A threshold- free summary
      index of prediction accuracy for censored time to event data. *Stat Med* 2018; **37**: 1671–81.

30   Steyerberg EW, Vickers AJ, Cook NR, *et al.* Assessing the Performance of Prediction
      Models: A Framework for Traditional and Novel Measures. *Epidemiology* 2010; **21**: 128–
      38.