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)

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**

Summary

Background Cancer treatment can cause gonadal impairment. Acute ovarian failure is defined as the permanent loss of ovarian function within 5 years of cancer diagnosis. We aimed to develop and validate risk prediction tools to provide accurate clinical guidance for paediatric patients with cancer.

Methods In this cohort study, prediction models of acute ovarian failure risk were developed using eligible female US and Canadian participants in the Childhood Cancer Survivor Study (CCSS) cohort and validated in the St Jude Lifetime Cohort (SJLIFE) Study. 5-year survivors from the CCSS cohort were included if they were at least 18 years old at their most recent follow-up and had complete treatment exposure and adequate menstrual history (including age at menarche, current menstrual status, age at last menstruation, and menopausal aetiology) information available. Participants in the SJLIFE cohort were at least 10-year survivors. Participants were excluded from the prediction analysis if they had an ovarian hormone deficiency, had missing exposure information, or had indeterminate ovarian status. The outcome of acute ovarian failure was defined as permanent loss of ovarian function within 5 years of cancer diagnosis or no menarche after cancer treatment by the age of 18 years. Logistic regression, random forest, and support vector machines were used as candidate methods to develop the risk prediction models in the CCSS cohort. Prediction performance was evaluated internally (in the CCSS cohort) and externally (in the SJLIFE cohort) using the areas under the receiver operating characteristic curve (AUC) and the precision-recall curve (average precision [AP; average positive predictive value]).

Findings Data from the CCSS cohort were collected for participants followed up between Nov 3, 1992, and Nov 25, 2016, and from the SJLIFE cohort for participants followed up between Oct 17, 2007, and April 16, 2012. Of 11 336 female CCSS participants, 5886 (51·9%) met all inclusion criteria for analysis. 1644 participants were identified from the SJLIFE cohort, of whom 875 (53·2%) were eligible for analysis. 353 (6·0%) of analysed CCSS participants and 50 (5·7%) of analysed SJLIFE participants had acute ovarian failure. The overall median follow-up for the CCSS cohort was 23·9 years (IQR 20·4–27·9), and for SJLIFE it was 23·9 years (19·0–30·0). The three candidate methods (logistic regression, random forest, and support vector machines) yielded similar results, and 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 haematopoietic stem-cell transplantation, cumulative alkylating drug dose, and an interaction between age at cancer diagnosis and haematopoietic stem-cell transplant. External validation of the model in the SJLIFE cohort 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. Based on these models, an online risk calculator has been developed for clinical use.

Interpretation Both acute ovarian failure risk prediction models performed well. The ovarian dose model is preferred if ovarian radiation dosimetry is available. The models, along with the online risk calculator, could help clinical discussions regarding the need for fertility preservation interventions in girls and young women newly diagnosed with cancer.

Funding Canadian Institutes of Health Research, Women and Children's Health Research Institute, National Cancer Institute, and American Lebanese Syrian Associated Charities.

Lancet Oncol 2020

Published Online
February 14, 2020
https://doi.org/10.1016/S1470-2045(19)30818-6

See Online/Comment
https://doi.org/10.1016/S1470-2045(20)30065-6

*Equally contributing senior authors

School of Public Health (R A Clark MSc, N K Vu PhD, Z Lu MEng, Y Yuan PhD) **and Faculty of Medicine and Dentistry** (T Motan MB ChB), **University of Alberta, Edmonton, AB, Canada; University of Pennsylvania Perelman School of Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA, USA** (S Mostoufi-Moab MD); **St Jude Children's Research Hospital, Memphis, TN, USA** (Y Yasui PhD, R J Brooke PhD, T M Gibson PhD, L L Robison PhD, W Chemaitilly MD, Prof M M Hudson MD, G T Armstrong MD); **Memorial Sloan Kettering Cancer Center, New York, NY, USA** (Prof C A Sklar MD); **Duke University School of Medicine, Durham, NC, USA** (Prof K C Oeffinger MD); **The University of Texas MD Anderson Cancer Center, Houston, TX, USA** (R M Howell PhD, S A Smith MPH); **and The Hospital for Sick Children, Toronto, ON, Canada** (Prof P C Nathan MD)

Correspondence to:
Dr Yan Yuan, School of Public Health, University of Alberta, Edmonton T6G 1C9, AB, Canada
yyuan@ualberta.ca

Introduction

Because of advancements in cancer treatment and supportive care, there are almost 500 000 survivors of childhood cancer in the USA,[1] and between 300 000 and 500 000 survivors of childhood cancer in Europe.[2,3] Survivors are at an increased risk of developing chronic health conditions from toxicities related to cancer treatment.[4] The cumulative burden of treatment-associated

Research in context

Evidence before this study

An increased risk of premature gonadal failure has been shown in survivors of paediatric cancer treated with chemotherapy and radiotherapy. 6% of female survivors of childhood cancer lose ovarian function within 5 years of treatment (acute ovarian failure), and an additional 9% have premature, non-surgical menopause before the age of 40 years. The timeframe between primary cancer diagnosis and treatment resulting in acute ovarian failure is short 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 acute ovarian failure and the associated risk factors in survivors of childhood cancer 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 acute ovarian failure as an independent condition without grouping patients in a broader premature menopause category. Although high-dose pelvic radiotherapy, haematopoietic stem-cell transplantation, and alkylating chemotherapy have been identified as risk factors associated with acute ovarian failure, clinicians do not have a tool that accurately estimates the risk of acute ovarian failure for individual paediatric patients with cancer at the time of cancer diagnosis. We did not find any study that aimed to develop risk estimates of acute ovarian failure for individual paediatric patients with cancer 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 acute ovarian failure in female survivors of childhood cancer. Although physicians are aware of the gonadotoxic treatment exposures with a high likelihood of causing acute ovarian failure, there are no available prediction tools to estimate the risk of acute ovarian failure for a given patient on the basis of a planned oncological treatment regimen. A precise risk estimate available to clinicians will guide informed discussions with patients and their families for time-sensitive interventions to preserve fertility function before 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 calculator for acute ovarian failure risk for clinicians to directly calculate each patient's risk for acute ovarian failure on the basis of their planned cancer treatment. The developed models performed well both internally and externally, highlighting the validity of the risk estimates and ensuring that clinical recommendations are provided with confidence.

Implications of all the available evidence

Since most patients with childhood cancer will become long-term survivors, the focus of cancer survivorship research has shifted toward maximising 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 before cancer treatment and the possible need for ovarian hormone replacement after completion of cancer therapy.

chronic health conditions is substantial, with survivors experiencing an average of 17 conditions by the age of 50 years (eg, endocrinal dysfunction, cardiac disease, and neurocognitive difficulties).[5] Impaired gonadal function is a common late effect of cancer therapy that can have a substantial impact on a survivor's quality of life.[6]

Girls are born with a finite supply of ovarian follicles that decline with age.[7] Cancer-directed therapies, such as radiotherapy and alkylating chemotherapy, can accelerate this decline, resulting in early cessation of ovarian endocrine and reproductive function.[7] Primary ovarian insufficiency is defined as compromised gonadal function before the age of 40 years.[8] Primary ovarian insufficiency can manifest as acute ovarian failure or premature menopause. Acute ovarian failure occurs when an individual permanently stops menstruating within 5 years of their cancer diagnosis, does not progress through puberty, or does not achieve menarche by 18 years of age following cancer treatment.[9] Previous investigations from the Childhood Cancer Survivor Study (CCSS) have shown a 6% prevalence of acute ovarian failure in female survivors of childhood cancer.[9] Premature menopause occurs when ovarian function is retained for at least 5 years following cancer diagnosis, but non-surgical menopause develops before the survivor reaches 40 years of age.[10] In the general population, the prevalence of premature, non-surgical menopause is approximately 1%,[11] whereas the cumulative incidence of premature menopause (excluding acute ovarian failure) reported in female survivors from the CCSS cohort is 9% by the age of 40 years.[10] Acute ovarian failure and premature menopause can restrict reproductive options, reduce quality of life, increase anxiety and depressive feelings, and increase the risk for serious morbidities including ischaemic heart disease, osteoporosis, and cognitive decline.[10,12–15]

Improved precision in predicting an individual's risk for developing primary ovarian insufficiency 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 because available fertility preservation technologies such as cryopreservation of ovarian tissue and oocytes are expensive and invasive. Ovarian tissue cryopreservation requires

surgery, the potential for future livebirths is poorly established, and there is concern that some malignancies might involve the ovary, precluding reimplantation of the tissue. The operative risk might be increased in children who are immunocompromised or have abnormal blood counts, which increases their risk for infection or bleeding. Oocyte harvest can only be offered to post-pubertal girls and women, and ovarian hyperstimulation can require up to 4–6 weeks before oocytes can be retrieved,[16] which impedes the use of harvesting before gonadotoxic treatment is initiated. No studies have addressed the safety or success rate of these procedures in children and adolescents.[17,18]

Although physicians are aware of the gonadotoxic treatment exposures that are likely to cause acute ovarian failure, no prediction tools are available to estimate the acute ovarian failure risk for a given patient on the basis of a planned oncological 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 acute ovarian failure at the time of cancer diagnosis to inform clinicians of the need for time-sensitive interventions. By offering fertility preservation procedures before treatment initiation to patients at a substantial risk of acute ovarian failure, the opportunity for future reproduction in this group can be maximised, and the risk of performing unnecessary procedures on patients at low risk of acute ovarian failure can be minimised.

Methods

Study design and participants

The CCSS, a multi-institutional longitudinal cohort study of 24 362 survivors of childhood cancer from the USA and Canada, was the primary data source for this study. Established in 1994, its cohort characteristics, eligibility criteria, and study design features have been documented elsewhere.[19,20] Briefly, the cohort includes 5-year survivors diagnosed before the age of 21 years with an eligible cancer type (leukaemia, CNS cancers, Hodgkin lymphoma, non-Hodgkin lymphoma, Wilms' tumour, neuroblastoma, soft-tissue sarcoma, or bone tumours) and treated at one of 31 participating institutions in the USA and Canada between Jan 1, 1970, and Dec 31, 1999. Survivors were eligible for the acute ovarian failure 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 the cause of menopause (surgical *vs* non-surgical) for those who were currently menopausal.[9] The ovarian status was deemed not assessable for participants who had been exposed to a pituitary radiation dose higher than 30 Gy, who had had a tumour in the hypothalamus or 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 data on radiation exposure and chemotherapy were also excluded.

The validation cohort was derived from the St Jude Lifetime Cohort (SJLIFE; NCT00760656), which was established in September, 2007, to medically assess the health status of survivors of childhood cancer treated at St Jude Children's Research Hospital (Memphis, TN, USA).[4] 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 10-year survivors.[4] Participants were excluded from the validation study if they had an ovarian hormone deficiency, were missing treatment exposure information, or if their ovarian status could not be determined.[8] Individuals who participated in both cohorts (CCSS and SJLIFE) were excluded from model development using CCSS and were 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 institution approved the CCSS protocol, the St Jude Children's Research Hospital approved the SJLIFE protocol, and written informed consent was obtained from all study participants.

Procedures

Self-reported demographic and outcome information between Nov 3, 1992, and Nov 25, 2016, was obtained from CCSS participants through baseline and follow-up questionnaires, and treatment exposure information was abstracted from medical and radiotherapy records, starting from Jan 1, 1970, to Dec 31, 2004.[20] For this analysis, radiation doses were estimated, including all therapy received within 5 years of primary cancer diagnosis, to the ovaries and two body regions that have the potential to overlap with the ovaries (abdomen and pelvis).[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.

Classification of ovarian status was assigned to CCSS participants using an established definition,[6] or manually by endocrinologists (SM-M and CAS) for ambiguous cases, on the basis of responses to menstrual history questions at baseline and follow-up questionnaires. Individuals who had had menarche before cancer treatment were diagnosed with acute ovarian failure if they did not resume menstruating within 5 years of cancer diagnosis. Individuals who had not had menarche were diagnosed with acute ovarian failure if menarche had not happened by the age of 18 years.[9] Age of 18 years was selected to avoid misdiagnosing participants with

acute ovarian failure who might not have had menarche by age 15 or 16 years because of cancer-related issues, such as poor nutrition and weight loss, but had menarche 1 or 2 years later. The remaining individuals in the study sample were classified as not having developed acute ovarian failure. Because acute ovarian failure is an early toxicity event occurring during or shortly after treatment completion, acute ovarian failure status was assessable at survey completion after survivors reached 18 years of age, and analytical methods for a binary outcome were used to analyse it. In survivors who reported taking an oral contraceptive pill, the survey differentiated between those still menstruating but taking a contraceptive pill for contraception purposes, those taking a contraceptive pill or hormone supplements to regulate their menstrual cycles, and those taking a contraceptive pill or hormone supplements as ovarian replacement therapy. Ovarian status classifications for SJLIFE participants were ascertained by an endocrinologist (WC) on the basis of questionnaire responses and hormone measurements.[22,23] The participants underwent clinical evaluations that considered data from self-reported history, previous medical records, and laboratory testing in the assessment of ovarian status. This assessment allowed us to evaluate the prediction algorithms for clinically verified classifications of ovarian status in this independent validation cohort.

Potential predictors for model development included age at cancer diagnosis, age at menarche, type of cancer at diagnosis, any exposure to chemotherapy, cumulative dose of alkylating drugs measured using the cyclophosphamide-equivalent dose,[24] haematopoietic stem-cell transplant, and ovarian radiation doses for an ovarian dose model. Because information about ovarian radiation dosage is not universally available, we also considered a prescribed dose model, in which 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 used to select the best model for each method.

See Online for appendix

Model performance was evaluated both internally and externally. To avoid over-optimistic estimates of performance during internal validation, the eligible CCSS data were randomly divided into a training set (75% of participants) and a test set (25% of participants) using a computer-generated random number sequence. A model was built on the training set using the variables selected in the model development stage, and the risk of acute ovarian failure was predicted using this model for participants in the corresponding test set. We repeated the randomisation process 100 times on 100 randomly split training and test sets and took the average of the predicted risk for each participant. Model performance was evaluated using the average predicted risk and the observed acute ovarian failure status.

Along with the continuous risk estimate, we wanted to provide users with an optional categorisation of the acute ovarian failure risk from the model. Predicted risks from the best models were used to stratify participants into categories of low (<5%), medium-low (5% to <20%), medium (20% to <50%), and high risk (≥50%). Risk value thresholds for the categories were defined by two endocrinologists (SM-M and CAS) and a paediatric oncologist (PCN) with expertise in childhood cancer survivorship, so that the groups represented reasonable and clinically meaningful categories for oncofertility discussions and decision making. A web-based calculator that predicts the risk of acute ovarian failure and assigns risk category on the basis of the ovarian dose model has been developed to facilitate the clinical use of our model. A similar calculator using prescribed doses (ie, pelvic or abdominal irradiation) is under development.

Outcomes

The primary outcome was presence or absence of acute ovarian failure, defined as the permanent loss of ovarian function within 5 years of cancer diagnosis or not having reached menarche by the age of 18 years following cancer treatment.

Statistical analysis

A detailed description of the statistical methods is available in the appendix (pp 1, 2). We developed candidate prediction algorithms using three popular methods for binary outcome analysis (logistic regression, random forest,[25] and support vector machines[26]), which allow the sensitivity of the data to the modelling method to be examined. 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 done 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 participants with acute ovarian failure from those without acute ovarian failure when presented with a pair of observations, one with acute ovarian failure and the other without. AUC values range from 0·5 to 1·0, with values closer to 1·0 indicating better performance. Precision-recall curves and area under the precision-recall curve, known as average precision (AP; or average positive predictive value) values, were used to measure the ability of the model to detect cases of acute ovarian failure from the entire population of female survivors of childhood cancer.[27–29] The positive predictive value represents the probability that a participant classified as having acute ovarian failure truly has acute ovarian failure. The AP value can be interpreted as the risk of acute ovarian failure for a participant whose predicted risk is greater than the risk estimate of a randomly

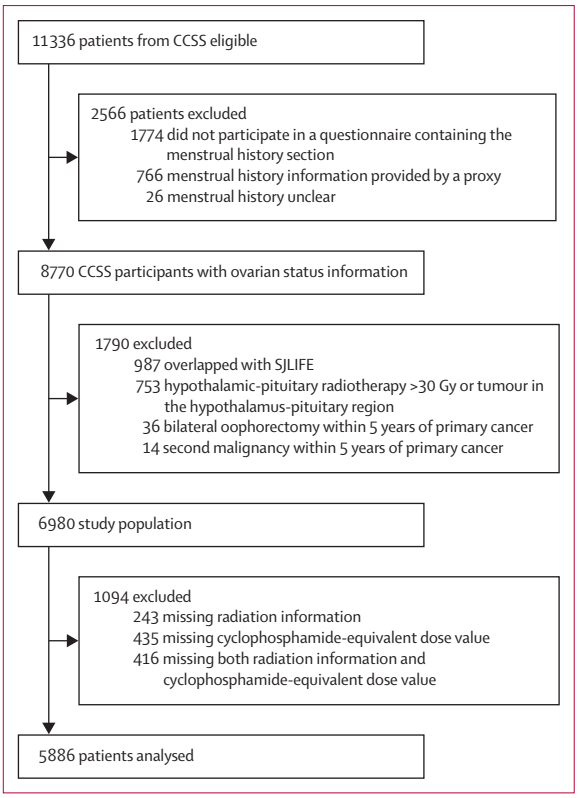

Figure 1: **CCSS study cohort flowchart for inclusion in model development**
The reasons for non-participation are unknown. CCSS=Childhood Cancer Survivor Study.

	CCSS cohort (n=5886)	SJLIFE cohort (n=875)
Median age at cancer diagnosis, years	7·3 (3·2–13·7)	6·7 (3·4–13·2)
Cancer diagnosis		
Leukaemia	1869 (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%)
CNS	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*	..	81 (9·3%)
Cumulative alkylating drug dose per CED value, g/m²		
None	3130 (53·2%)	363 (41·5%)
<4	801 (13·6%)	75 (8·6%)
4 to <8	743 (12·6%)	163 (18·6%)
≥8	1212 (20·6%)	274 (31·3%)
Minimum ovarian radiation dose, Gy		
None	3174 (53·9%)	710 (81·1%)
<10	2383 (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	3173 (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	3173 (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%)
Haematopoietic stem-cell transplant		
Yes	217 (3·7%)	18 (2·1%)
No	5669 (96·3%)	857 (97·9%)

Data are median (IQR) or n (%). CCSS=Childhood Cancer Survivor Study. SJLIFE=St Jude Lifetime Cohort. CED=cyclophosphamide-equivalence dose.
*The SJLIFE cohort has less restrictive inclusion criteria than the CCSS cohort, and 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.

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

selected female survivor with acute ovarian failure. AP values range from the value of the event rate in the population to 1·0 (ie, the highest possible value), and when comparing candidate models of specific populations, larger values indicate superior performance. Scaled Brier scores for overall model performance evaluation (larger values are preferred when comparing models)[30] were calculated, and calibration curves were used to visually inspect the alignment of the predicted risk of acute ovarian failure with the observed risk of acute ovarian failure.[30]

Role of the funding source
The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to CCSS data, and YYa, LLR, MMH, and GTA had full access to CCSS and SJLIFE data. The corresponding author had final responsibility for the decision to submit for publication.

Results
Data were collected for participants in the CCSS cohort followed up between Nov 3, 1992, and Nov 25, 2016, and for participants in the SJLIFE cohort followed up between Oct 17, 2007, and April 16, 2012. Of the total 11 336 female CCSS participants identified, 5450 (48·1%) were excluded (figure 1). 8770 (77·4%) participants had ovarian status information, and 5886 (51·9%) participants met all eligibility criteria and were included in the analysis. Of the 1644 total female survivors in the SJLIFE cohort, 656 (39·9%) did not respond, declined to participate, only consented to the survey but not the clinical evaluation, or did not complete the clinical visit and were excluded. 67 (4·1%) were excluded due to hormone

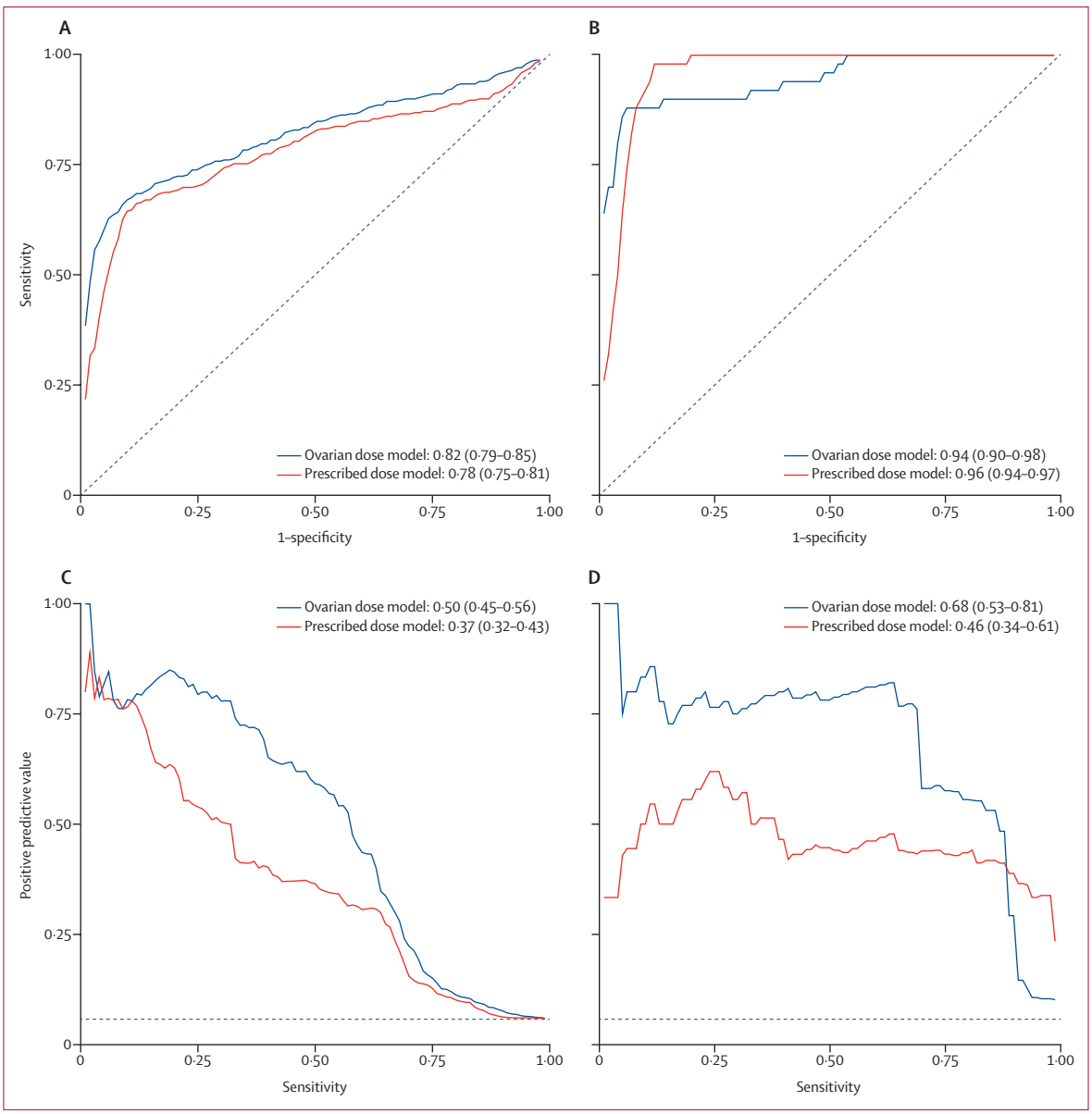

Figure 2: **ROC and precision-recall curves for the best models**
ROC curves and AUC values with 95% CIs for the ovarian dose model (blue) and the prescribed dose model (red) from the CCSS cohort (A) and the SJLIFE cohort (B). Precision-recall curves and AP values with 95% CIs for the ovarian dose model (blue) and the prescribed dose model (red) from the CCSS cohort (C) and the SJLIFE cohort (D). AP=average precision. AUC=area under the ROC curve. CCSS=Childhood Cancer Survivor Study. ROC=receiver operating characteristic. SJLIFE=St Jude Lifetime Cohort.

deficiency, undetermined origin of ovarian insufficiency, or undetermined ovarian status, and an additional 46 (2·8%) survivors were excluded because of missing radiation dose data (n=38) or a subsequent malignancy (n=8). Thus, a total of 875 survivors in SJLIFE were eligible and included in the analysis

Demographic, diagnostic, and treatment characteristics of the 5886 eligible CCSS survivors and the 875 eligible SJLIFE survivors are presented in table 1. Almost a third (1869 [31·8%] of 5886) of the CCSS survivors had been diagnosed with leukaemia, and 217 (3·7%) had a haematopoietic stem-cell transplant, whereas 378 (43·2%) survivors from SJLIFE had been diagnosed with leukaemia and 18 (2·1%) had a haematopoietic stem-cell transplant. Survivors in the SJLIFE cohort were exposed, on average, to higher doses of alkylating drugs than survivors in CCSS, but fewer survivors (165 [18·9%] *vs* 2712 [46·1%]) received radiation to at least one of the ovaries. Survivors in the CCSS analysis cohort had an overall median follow-up of 23·9 years (IQR 20·4–27·9) and a total follow-up of 142738·8 person-years. For survivors in the SJLIFE analysis cohort, the median

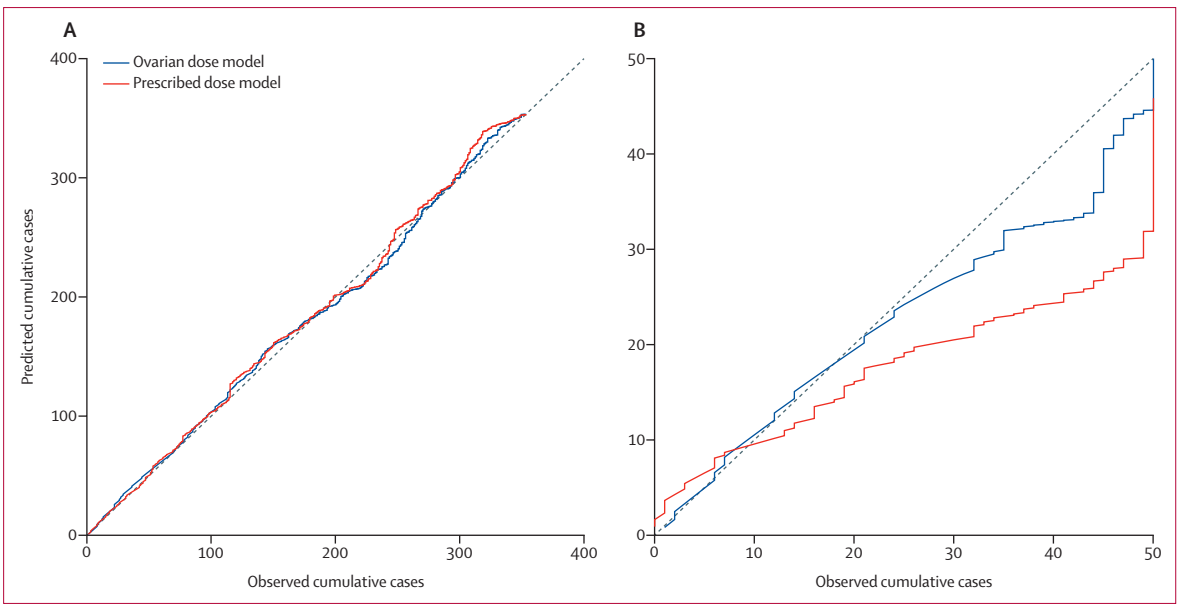

Figure 3: **Calibration curves for the best models**
Calibration curves for the ovarian dose model (blue) and the prescribed dose model (red) in the CCSS cohort (A) and the SJLIFE cohort (B). CCSS=Childhood Cancer Survivor Study. SJLIFE=St Jude Lifetime Cohort.

follow-up was 23·9 years (19·0–30·0) and the total follow-up was 21492·2 person-years.

Because acute ovarian failure is an early toxicity event, every survivor's acute ovarian failure status in these two cohorts was ascertained after either 18 years of age or 5 years after cancer diagnosis, depending on her menarcheal status at cancer diagnosis. Thus, the effective follow-up time for determining acute ovarian failure was a median of 5·0 years (IQR 5·0–9·1) and a total follow-up of 42335·1 person-years in the CCSS analysis cohort and 5·0 years (5·0–8·8) and 6150·7 person-years in the SJLIFE analysis cohort. 353 (6·0%) of 5886 survivors in the CCSS cohort were diagnosed with acute ovarian failure. A similar prevalence was observed for the SJLIFE survivors; 50 (5·7%) of 875 survivors in this cohort were diagnosed with acute ovarian failure. Of the survivors in the CCSS cohort, 84 (1·4%) had started taking oral contraceptive pills within 5 years of their cancer diagnosis and were classified as not having acute ovarian failure after detailed review of their survey responses by two endocrinologists (SM-M and CAS). Use of oral contraceptives was not the only consideration for ovarian status determination in SJLIFE as patients were also assessed on the basis of clinical hormonal measurements, medical records, and self-report.

The best models from the three candidate methods (logistic regression, random forest, and support vector machines) yielded similar results (appendix p 6). Therefore, the logistic regression models were selected because of the transparency and interpretability of logistic regression compared with the other methods. Variables and their estimated coefficients for the best ovarian dose model and prescribed dose model are shown in the appendix (pp 4, 5). Common predictors for both models were history of haematopoietic stem-cell transplantation, cumulative alkylating drug dose, and an interaction between age at cancer diagnosis and haematopoietic stem-cell transplant (appendix pp 4, 5).

The ROC curves of the best prediction models are shown in figure 2A and 2B. When the CCSS test sets were used for internal evaluation, the best ovarian dose model's AUC value was 0·82 (95% CI 0·79–0·85) and the best prescribed dose model's AUC value was 0·78 (0·74–0·81; figure 2A). The AUC values for both models increased when the SJLIFE cohort was used as an independent validation set (figure 2B). The ovarian dose model's AUC value was 0·94 (0·90–0·98) and the prescribed dose model's AUC value was 0·96 (0·94–0·97).

The precision-recall curves for the best models are shown in figures 2C and 2D. For internal validation using the CCSS test sets, the AP value of the precision-recall curve was 0·50 (95% CI 0·45–0·56) for the ovarian dose model and 0·37 (0·32–0·43) for the prescribed dose model (figure 2C). Similar to the AUC results, the AP values increased when externally validated in SJLIFE, with an AP of 0·68 (0·53–0·81) for the ovarian dose model and 0·46 (0·34–0·61) for the prescribed dose model (figure 2D).

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 was 20·0% for the CCSS cohort and 22·6% for the SJLIFE cohort.

The calibration curves for the CCSS cohort closely follow the diagonal for both the ovarian dose and prescribed dose model, indicating good alignment

For the **acute ovarian risk prediction calculator** see https://ccss.stjude.org/aofcalc

	Survivors	Survivors with acute ovarian failure
Best ovarian dose model		
Low (<5%)		
CCSS	5130/5886 (87·2%)	119/5130 (2·3%)
SJLIFE	796/875 (91·1%)	8/796 (1·0%)
Medium-low (5% to <20%)		
CCSS	429/5886 (7·3%)	47/429 (11·0%)
SJLIFE	34/875 (3·9%)	8/34 (23·5%)
Medium (20% to <50%)		
CCSS	145/5886 (2·5%)	55/145 (37·9%)
SJLIFE	8/875 (0·9%)	4/8 (50·0%)
High (≥50%)		
CCSS	182/5886 (3·1%)	132/182 (72·5%)
SJLIFE	37/875 (4·2%)	30/37 (81·1%)
Best prescribed dose model		
Low (<5%)		
CCSS	4898/5886 (83·2%)	117/4898 (2·4%)
SJLIFE	709/875 (81·0%)	1/709 (0·1%)
Medium-low (5% to <20%)		
CCSS	515/5886 (87·5%)	62/515 (12·0%)
SJLIFE	114/875 (13·0%)	26/114 (22·8%)
Medium (20% to <50%)		
CCSS	376/5886 (6·4%)	113/376 (30·1%)
SJLIFE	39/875 (4·5%)	17/39 (43·6%)
High (≥50%)		
CCSS	97/5886 (1·6%)	61/97 (62·9%)
SJLIFE	13/875 (1·5%)	6/13 (46·2%)

Data are n/N (%). CCSS=Childhood Cancer Survivor Study. SJLIFE=St Jude Lifetime Cohort.

Table 2: **Acute ovarian failure risk categories and prevalence for each cohort as predicted by the best ovarian dose model and the best prescribed dose model**

between the observed and predicted risk (figure 3A). For the SJLIFE cohort, although the calibration curve for the ovarian dose model mostly follows a diagonal line, a slight deviation toward its end suggests that the model is not well calibrated at the low end of the risk. The calibration curve for the prescribed dose model deviates from the diagonal line more than that of the ovarian dose model, indicating inferior reliability of the risk estimates (figure 3B).

The ovarian dose model distinguished between low-risk and high-risk survivors with more certainty than the prescribed dose model, as seen by the higher number of survivors classified into the high (≥50%) and low (<5%) risk groups (table 2). Specifically, using the ovarian dose model, 5130 (87·2%) of 5886 participants in the CCSS cohort were estimated to be at low risk (of whom 119 [2·3%] developed acute ovarian failure), whereas the prescribed dose model estimated 4898 (83·2%) of 5886 individuals to be at low risk (of whom 117 [2·4%] developed acute ovarian failure). At the high end of the predicted risk spectrum, the ovarian dose model classified 182 (3·1%) individuals as high risk (of whom 132 [72·5%] had acute ovarian failure), whereas the prescribed dose model predicted only 97 (1·6%) individuals as high-risk, with 61 (62·9%) of them having developed acute ovarian failure. A cross-table of the risk estimates from the prescribed dose and ovarian dose models on the basis of the CCSS cohort and a detailed comparison is included in the appendix (pp 2, 5).

A web-based application of the acute ovarian failure models has been developed to calculate risk of acute ovarian failure in an individual patient. To use this application, clinicians input the patient's age at cancer diagnosis and proposed treatment exposures (whether they will undergo a haematopoietic stem-cell transplant, the total body irradiation dose, any additional abdominal and pelvic or ovarian radiation doses, and either the cyclophosphamide-equivalent dose value or specific alkylating drug doses). A continuous risk estimate of acute ovarian failure and corresponding risk category is calculated on the basis of the radiation information provided.

Discussion

With AUC values ranging 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 acute ovarian failure that provide a high level of confidence appropriate for use in a clinical setting. Because AP values larger than the population event rate imply superior predictive ability to detect cases of acute ovarian failure, values for the prescribed dose and ovarian dose models that range between 0·37 and 0·68 indicate strong predictive power for detecting acute ovarian failure compared with the event rate of 0·06 in both cohorts. Given the short time interval between cancer diagnosis, treatment, and the subsequent development of acute ovarian failure, 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 strong performance of the models in the external SJLIFE cohort supports the concept that our prediction algorithms are generalisable. The main outcome for our risk prediction model was acute ovarian failure, which was classified using self-reported menstrual history information provided in CCSS questionnaires. However, the ovarian status of CCSS participants was not verified clinically and was therefore 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 concentrations, permitting a more precise ascertainment of ovarian status than in the CCSS cohort.[8,22,23] Although outcome data in CCSS, which we

used to develop our model, have a higher potential for misclassification than those in SJLIFE, 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 doses of alkylating drugs than survivors in CCSS, but fewer survivors received radiotherapy to the ovaries. This difference might reflect differences in the distribution of diagnoses between the two cohorts, in addition to the fact that some participants in SJLIFE were treated more recently when, in an effort to improve late health outcomes, there have been attempts to reduce radiation exposure by increasing the use of some chemotherapy drugs (eg, alkylating drugs and anthracyclines). 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 participants into the low-risk and high-risk categories, with a smaller number of participants categorised into the intermediate-risk categories, which have the greatest uncertainty of all categories regarding the necessity of intervention. However, to use the ovarian dose model, the estimated radiation dose to the ovaries is required. This information might not be available to all health-care providers. By 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 acute ovarian failure risk prediction.

Validation of the risk models in a prospective cohort would further establish the validity of the online risk prediction tool that we developed. This is not currently planned, but since the risk prediction tool is freely available online, we 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 risk of acute ovarian failure with patients and families. Since some patients considered at low or moderate risk for acute ovarian failure at first presentation of cancer might nonetheless develop acute ovarian failure, whereas 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 acute ovarian failure 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 performed well at predicting acute ovarian failure cases in the population of survivors of childhood cancer, there are limitations. Cases of acute ovarian failure might be masked in female survivors who are taking oral contraceptives or other hormone medications that result in persistent menstruation and might result in an underestimate of the prevalence of acute ovarian failure. 84 survivors in the CCSS cohort who were classified as not having acute ovarian failure started taking oral contraceptive pills within 5 years of their cancer diagnosis and were thus at risk for misclassification, but detailed review by the two study endocrinologists minimised this risk. The reliance on self-reported menstrual history and possible misclassification of participants at risk for both primary and central hypogonadism represent additional limitations, but the fact that the models perform well in the SJLIFE cohort, which uses clinically verified hypogonadism and ovarian status in amenorrhoeic survivors younger than 40 years, provides reassurance.

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

A complete case analysis approach, rather than multiple imputation, was used in our study and can thus result in bias and inefficiency. However, the excellent prediction performance of our models on an external cohort, and the 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 participants 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 effect of such therapies on ovarian function emerges.

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

Contributors

YYu, PCN, KCO, YYa, SM-M, CAS, and TM developed the concept for the study. YYa, GTA, KCO, TMG, LLR, MMH, and YYu designed the

study. YYa, RAC, GTA, SM-M, CAS, TMG, RMH, SAS, LLR, WC, MMH, and YYu collected the data. RAC, NKV, RJB, ZL, YYu, and YYa prepared and analysed the data. NKV, ZL, RJB, and YYu created the figures. RAC, NKV, ZL, YYu, YYa, PCN, SM-M, CAS, and WC interpreted the results. RAC, SM-M, PCN, YYu, YYa, CAS, RMH, SAS, RJB, MMH, GTA, WC, LLR, TM, NKV, TMG, KCO, and ZL drafted and revised the manuscript.

Declaration of interests

CAS has received speaker's fees and travel grants from Novo Nordisk and royalties from UpToDate. RMH and SAS from the MD Anderson Late Effects Group have a subcontract with St Jude Hospital Research Center for CCSS dosimetry and a contract from Radiation Epidemiology Branch of the National Cancer Institute to perform dosimetry on various studies. RAC has a graduate studentship awarded by Women and Children's Health Research Institute for this study. YYu has a grant from the Canadian Institutes for Health Research that supported some aspect of this study. GTA and YYa (R01 CA216354) have grants from the National Cancer Institute that supported some aspect of this study. TMG has a grant from the National Institutes of Health that supported some aspect of this study. All other authors declare that they have no competing interests.

Acknowledgments

We thank the Canadian Institutes for Health Research for providing funding (FRN 148693). The study data is from the Childhood Cancer Survivor Study (grant number U24 CA55727), the St Jude Lifetime Cohort Study (supported by National Cancer Institute grant U01CA195547) Cancer Center Support CORE grant CA21765; the American Lebanese Syrian Associated Charities (Memphis, TN, USA). We thank the CCSS and SJLIFE study participants and their families for providing the time and effort for participation; Wendy Leisenring and the statistical team at the Fred Hutchinson Cancer Research Center (Seattle, WA, USA) for their help in obtaining and comprehending the CCSS dataset; and the internet team at St Jude Children's Research Hospital for the development of the web application of the risk prediction models.

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:** 3240–47.

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:** 1723–28.

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 Coyne K, Purdy M, O'Leary K, Yaklic JL, Lindheim SR, Appiah LA. Challenges and considerations in optimizing ovarian stimulation protocols in oncofertility patients. *Front Public Health* 2014; **2:** 246.

17 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.

18 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. **54:** 2200–09.

19 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.

20 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.

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:** 855–59.

28 Yuan Y, Su W, Zhu M. Threshold-free measures for assessing the performance of medical screening tests. *Front Public Health* 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.