CCSS | Analysis Concept Proposal

Study Title

Quantifying the individual absolute risk of premature ovarian failure in female survivors of childhood cancer

Working Group

Primary: Biostatistics/Epidemiology
Secondary: Cancer control, Chronic disease

Investigators

Name	Contact Information
Dr. Yan Yuan	yyuan@ualberta.ca
Dr. Paul Nathan	paul.nathan@sickkids.ca
Dr. Yutaka Yasui	yutaka.yasui@stjude.org
Dr. Tarek Motan	tmotan@ualberta.ca
Dr. Melanie Barwick	melanie.barwick@sickkids.ca
Dr. Charles Sklar	sklarc@mskcc.org
Dr. Sogol Mostoufi-Moab	moab@email.chop.edu
Dr. Gregory Armstrong	greg.armstrong@stjude.org
Dr. Rebecca Maureen Howell	rhowell@mdanderson.org
Dr. Todd M. Gibson	todd.gibson@stjude.org
Dr. Kevin Oeffinger	oeffingk@mskcc.org
Dr. Melissa M. Hudson	melissa.hudson@stjude.org
Rebecca Clark (MSc Candidate)	raclark@ualberta.ca

Background and Rationale

There are currently more than 30,000 survivors of childhood cancer in Canada[1], and over 400,000 survivors in each of the US and Europe[2,3]. This is a testimony to the effectiveness of modern multimodal cancer therapies that improved the less than 30% survival in 1960s to 80% today. However, improved survival comes at a cost; healthcare providers have increasingly recognized a number of late sequelae of childhood cancer and its treatments. A frequent and significant late effect of cancer therapy is compromised reproductive function in both males and females[4,5]. Female survivors with non-surgical loss of ovarian function are classified into two subtypes, acute and late

onset, depending on the timing of ovarian failure[6]. *Acute ovarian failure* (AOF) refers to loss of ovarian function during or shortly after cancer treatment, typically within 5 years of cancer diagnosis. Late onset ovarian failure, also known as *premature menopause* (PM), occurs when patients retain ovarian function after the completion of cancer treatment but cease to menstruate prematurely before age 40 years. Previous studies have demonstrated that: 1) female childhood cancer survivors are 10.5 times more likely to develop PM when compared to their female non-cancer siblings[7]; 2) the estimated prevalence of AOF and PM are 6.3% and 9.0%, respectively in a long-term childhood cancer survivor cohort[8,9]; and 3) PM occurs as early as age 20-30 years[10].

Factors that strongly influence the risk of AOF and PM have been reported in both childhood and adult cancer patients[5,8,10,11]. Risk factors for AOF include dose of radiation to the ovaries, dose and type of alkylating agents, older age at diagnosis, menarche status at treatment, and stem cell transplantation[6,8,11,12]. Significant interactions between age at diagnosis with radiation to the ovaries and with cyclophosphamide have been reported[8]. For PM, known risk factors include older age at diagnosis, treatment with cyclophosphamide equivalent dose (CED[13]) $\geq$ 6,000 mg/m^2 and ovarian radiation[7].

In healthy women, fertility decreases drastically around 10 years prior to menopause[12]. Thus, PM greatly reduces a survivor's reproductive window. Due to this concern, multiple national and international organizations have established guidelines for fertility consultation for cancer patients and survivors[14]. Fertility preservation technologies, such as ovarian tissue and oocyte cryopreservation, are available but expensive. Furthermore, ovarian tissue cryopreservation involves surgery and is still considered experimental, while the oocyte cryopreservation procedure can only be used post-puberty. Thus, it is imperative that clinicians only offer these procedures to patients and survivors who are at significant risk of premature ovarian failure.

Although risk factors for ovarian failure have been well documented, reliable tools for accurately quantifying future fertility potential are lacking, limiting clinicians' ability to counsel patients and survivors based on a robust estimate of their risk. A newly developed tool will allow physicians to reassure patients and survivors who are at low risk of premature ovarian failure, allowing the avoidance of invasive and expensive treatments, thus benefiting the health care system and positively impacting the quality of life of patients and survivors. In contrast, clinicians will be able to thoroughly discuss different fertility preservation options with patients and survivors at high risk of premature ovarian failure. We propose to develop, validate and disseminate a novel and clinically useful tool that predicts the absolute risk of AOF and PM for individual childhood cancer patients and survivors. Our ultimate goal is to improve the quality of life of childhood cancer survivors in their adulthood.

Specific Aims

1. To quantify the complex relationship between age at diagnosis, cancer diagnosis and cancer treatment with the risk of premature ovarian failure (both AOF and PM), using stratified analysis and tree-based methods. Subsequently, standard and modern statistical methods will be used to build risk prediction models, followed by the selection of the best model using time-specific and overall accuracy measures;

2. To assess the accuracy of all risk prediction models in independent external validation cohorts;

Aims 3 and 4 represent knowledge translation elements of this project. We have included them here to provide a comprehensive view of the final goal and a complete picture of the overall project. These two aims are not part of the analysis and further CCSS data will not be required in order for their completion.

3. To develop simple risk scores for clinical use in collaboration with knowledge users (clinicians);

4. To create a web-based risk prediction calculator and a mobile app for knowledge dissemination in collaboration with knowledge translation experts.

Analysis Framework

(a) Outcomes of Interest

Premature Ovarian Failure (POF): Survivors are considered to have POF if they reported never menstruating by age 18 or spontaneous amenorrhea before age 40. Two subtypes are identified depending on the timing of the POF: *Acute Ovarian Failure* (AOF) if the patient never menstruated by age 18 or if she had OF within 5 years of cancer diagnosis; Otherwise, the POF is classified as *Premature Menopause* (PM), if the patient retained normal ovarian function following treatment completion but ceased to menstruate by age 40.

(b) Subject Population

- We will use data from the female population of the CCSS cohort of adult survivors (original and expansion cohort, n = 11 371) to establish the relationship between the identified covariates and to build the risk prediction models as defined in *Aim 1*.

- The data will be partitioned into training (75%) and test (25%) data sets
- The 25% test data set will be used for internal validation
- Data sets acquired from the St. Jude Lifetime Cohort study (SJLIFE; including additional biomarkers LH/FSH) and the DCOG-LATER-VEVO study (DCOG) for external validation as outlined in *Aim 2.* A letter of support from SJLIFE is enclosed and endorsement from the principal investigator of the DCOG study, Dr. Marleen van den Berg, was received (official approval by the DCOG board has been requested)

(c) Exploratory Variables

Baseline variables

- Patient Identification number
- Date of Birth
- Date of Diagnosis
- Initial Cancer Diagnosis (type)
    - Acute lymphoblastic leukemia, other leukemia, Hodgkin lymphoma, other lymphoma, brain tumour, neuroblastoma, kidney tumour, soft tissue sarcoma, bone tumour, other neoplasm
- Ethnicity
    - White, Black, American Indian or Alaskan Native, Asian or Pacific Islander
    - Hispanic (Yes/No)
- Height (at times of cancer treatment (if available) and each questionnaire completion)
- Weight (at times of cancer treatment (if available) and each questionnaire completion)

Treatment exposure

- Chemotherapy
    - Cyclophosphamide exposure (Yes/No) + Age at exposure
    - Procarbazine exposure (Yes/No) + Age at exposure
    - Number of alkylating agents in total
    - Dosage of each alkylating agent administered (if available, total dosage and dosage normalized to standard body surface area)
    - Cyclophosphamide Equivalent Dose (CED)[13]

- Radiation
    - Maximum prescribed tumour dose to the abdomen body region + age at exposure
    - Average dose to right and left ovaries + age at exposure

    - Maximum prescribed tumour dose to the pelvic body region + age at exposure
    - Average dose to pituitary gland + age at exposure (we may exclude individuals with doses $\geq$ 30 Gy from the modelling work, to be determined at project meeting)
    - Maximum prescribed total body irradiation dose + age at exposure

- Stem Cell Transplantation (Yes/No)

Smoking status

- Have you smoked at least 100 cigarettes in your entire life?
- On average how many a day do/did you smoke?
- How many years, in total, have you smoked?
- Current user (Yes/No)
- Former user (Yes/No)

d) Outcome variable

Menstrual History

- Did you need medication to go into puberty?
- Have you ever taken female hormones, including birth control pills (oral contraceptive) to have your period?
- Have you ever had a menstrual period + age at first occurrence?
- Are you currently experiencing menstrual periods?
- If no, what was the age at your last menstrual period?
- PM or AOF diagnosis (if available from CCSS database)

Modeling Approaches

Using CCSS data, we will use both classic and modern statistical modeling methods to develop the risk prediction models for AOF and PM as outlined in *Aim 1*. The classic models are used for benchmarking model performance.

Models for AOF: AOF occurs within 5 years of cancer diagnosis, for which a *prevalence analysis* is appropriate in the 5-year survivor cohort of CCSS. We will use two classification methods to develop the prevalence prediction model for AOF among 5-year survivors, namely classic logistic regression (the benchmark) and the modern statistical learning algorithm random forest (RF)[16] whose prediction accuracy outperformed logistic regression in many applications[17-19].

Models for PM: As a late effect of cancer treatments, PM may occur anytime during survivors' young adulthood post the 5-year survival milestone. We are interested in predicting PM *incidence over time*. The following four models for time to PM will be considered: a) the Cox proportional hazards (PH) model[20]; b) the accelerated failure time (AFT) model[21]; c) the time-specific logistic regression (TLR) model[22]; and d) an extension of the RF which handles time-to-event outcomes[23].

We propose to employ four models for empirical comparisons for various theoretical considerations. The PH model is the classic model for time-to-event outcomes in clinical research and was used for developing the Framingham Risk Score (FRS); therefore we will use it to set the benchmark. The AFT model is consistent with etiological hypothesis of PM where the treatment exposure (predictor in the model) speeds up the natural decline of ovarian reserve, thus "accelerating failure time". Unlike the AFT and PH models which model the entire event time with certain assumptions, the TLR directly models the risk of PM at a specific time, which aligns with our goal of absolute risk prediction within a specific time frame. This type of model extends the logistic regression by using inverse probability weights[24] to accommodate the censoring of event time. Lastly, the extended RF is advantageous in incorporating possible high-order interactions between predictors and leveraging the power of model ensemble. All models will take competing risks into consideration. It is not possible to know a priori which model will produce more accurate absolute risk in this particular application.

Model Development: CCSS data will be partitioned into training (75%) and test (25%) data sets. We will use stratified sampling to ensure that the proportions of AOF and PM cases in the two data sets are similar. Prediction models will be developed on the training data. The effects of continuous predictors, such as age at diagnosis, dosage of radiotherapy, and dosage of chemotherapy may be non-linear on the risk of ovarian failure. We will use graphical tools to visualize each of their effects to get the correct functional forms. We will utilize current knowledge of the interaction between risk factors in the development of accurate prediction model for each outcome. In addition, with clinical input, we will examine plausible interactions between risk factors by stratified analysis and the tree-based method[25]. To construct a final model for the logistic, PH, AFT, and TLR models, we will incorporate clinical consideration and use a modern statistical variable selection method (LASSO) that has been shown to improve prediction accuracy[26]. Both RF and its extension are data driven learning algorithms with built-in cross-validation and model averaging functionality. We will experiment with different tuning parameters to improve prediction accuracy for these two models.

Accuracy Assessment*:* We will conduct comprehensive model accuracy assessment for the <u>two</u> final models for AOF and the <u>four</u> final models for PM. For AOF, the predictive

accuracy of the logistic and RF models will be assessed with the Average Positive predictive value (AP)[27,28], and the discrimination will be assessed with Area Under Receiver Operating Characteristic Curve (AUC)[29]. For PM, the overall performance will be evaluated with calibration plots and Integrated Brier Score[30,31]. The time-specific predictive accuracy and discrimination will be assessed with AP_t [32] and AUC_t [33], respectively. Among the competing models, the project team will select one prediction model for each outcome, AOF and PM respectively, guided by accuracy measures and clinical considerations. External data will be requested from the St. Jude Lifetime Cohort study (SJLIFE) and the DCOG-LATER-VEVO study (DCOG) in order to assess the performance of the developed models through external validation as outlined in *Aim 2*.

Knowledge Translation (to be carried out after the completion of the analyses proposed here)

Aims 3 and 4 involve the creation of prediction tools for clinical use. These objectives will be undertaken following the development of *Aims 1 and 2*, the risk prediction models and subsequent accuracy assessment. These two *Aims* will not require the use of additional CCSS data. As the clinical utility of a prediction tool depends on its accuracy and simplicity, we plan to reduce the two final prediction models developed and validated under *Aims 1 and 2* (one for AOF and one for PM) to two simple risk scores, similar to the Framingham Risk Score. In addition, we will develop a user-friendly risk prediction calculator available through webpage and mobile app, for ease of use by clinicians and patients. The risk calculator will be designed for accurate risk prediction, as per our tested model, accompanied by clear interpretation of the predicted risk for end users.

Tables (for *Aims 1 and 2*)

Table 1: Comparing characteristics in the training cohort vs. test cohorts

	CCSS Training Data Set		**CCSS Test Data Set**		**SJLIFE**		**DCOG**	
Characteristic	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Age at Diagnosis								
0 – 4								
5 – 9								
10 – 14								
15 – 20								
Cancer Diagnosis								
Acute lymphoblastic leukemia								
Other leukemia								
Hodgkin lymphoma								
Brain tumour								
Neuroblastoma								
Kidney tumour								
Soft tissue sarcoma								
Bone tumour								
Other neoplasm								
Ethnicity								
White								
Black								
American Indian or Alaskan Native								
Asian or Pacific Islander								
Hispanic?								
Yes								
No								
Cyclophosphamide Exposure								
Yes								
No								
Unsure								
Age at Exposure								
<13								
13 - 20								
Unknown								

Procarbazine Exposure Yes No Unsure								
CED (mg/m^2) 0 > 0 – 3999 4000 – 5999 6000 – 7999 8000 +								
Abdominal/Pelvic RT Dose <5 Gy 5 - 10 Gy 11 - 20 Gy >20 Gy								
RT to HPO Axis Dose <5 Gy 5 - 10 Gy 11 - 20 Gy >20 Gy								
Total Body Irradiation Dose <12 Gy 12 Gy >12 Gy								
Stem Cell Transplant Yes No								
Smoking Status Current Former Never								
Smoking Duration								
Average quantity of cigarettes per day								

Outcome	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Incidence of AOF								
Cumulative Incidence of PM By age 30 By age 40								
Cumulative Incidence of competing risk events (death and hysterectomy) By age 30 By age 40								

Table 2: Model performance for risk of AOF

	Logistic Regression (benchmark)				**Random Forest Model**			
	Training	**Internal Validation**	**SJLIFE cohort**	**DCOG cohort**	**Training**	**Internal Validation**	**SJLIFE cohort**	**DCOG cohort**
Prediction Accuracy (AP)								
Discrimination(AUC)								

AP: Average Positive predictive value, AUC: Area Under Receiver Operating Characteristic Curve

Table 3: Model performance for risk of PM

	Cox Proportional Hazards (benchmark)				Accelerated Failure Time Model			
	Training	**Internal Validation**	**SJLIFE cohort**	**DCOG cohort**	**Training**	**Internal Validation**	**SJLIFE cohort**	**DCOG cohort**
Model Calibration (Calibration plots)								
Prediction Accuracy (AP_t)								
Discrimination (AUC_t)								
Overall Performance (Integrated Brier Score)								

AP_t: Average Positive predictive value at time t, AUC_t: Area Under Receiver Operating Characteristic Curve at time t

	Time-specific Logistic Regression Model				Random Forest Extension Model			
	Training	**Internal Validation**	**SJLIFE cohort**	**DCOG cohort**	**Training**	**Internal Validation**	**SJLIFE cohort**	**DCOG cohort**
Model Calibration (Calibration plots)								
Prediction Accuracy (AP_t)								
Discrimination (AUC_t)								
Overall Performance (Integrated Brier Score)								

AP_t: Average Positive predictive value at time t, AUC_t: Area Under Receiver Operating Characteristic Curve at time t

References

1. CIHR Institute of Cancer Research. The Late Effects of Childhood Cancer Treatments Initiative. 2012-04-12; At: http://www.cihr-irsc.gc.ca/e/documents/icr_late_effects_en.pdf.

2. Howlader N, Noone A, Krapcho M, et al. SEER cancer statistics review,1975-2012. Natl Cancer Inst 2015.

3. Winther J, Kenborg L, Byrne J, et al. Childhood cancer survivor cohorts in Europe. Acta Oncol 2015;54(5):655-668.

4. Howell S, Shalet S. Gonadal damage from chemotherapy and radiotherapy. Endocrinol. Metab Clin North Am 1998;27(4):927-943.

5. Letourneau J, Ebbel E, Katz P, et al. Acute ovarian failure underestimates age specific reproductive impairment for young women undergoing chemotherapy for cancer. Cancer 2012;118(7):1933-1939.

6. Sklar C. Maintenance of ovarian function and risk of premature menopause related to cancer treatment. J Natl Cancer Inst Monogr 2005;(34):25-27.

7. Sklar C. Memorial Sloan Kettering Cancer Center. 2016; Personal Communication.

8. Chemaitilly W, Mertens A, Mitby P, et al. Acute ovarian failure in the childhood cancer survivor study. J Clin Endocrinol Metab 2006;91(5):1723-1728.

9. Armstrong G. Childhood cancer survivor study. 2016; Personal Communication.

10. Sklar C, Mertens A, Mitby P, et al. Premature menopause in survivors of childhood cancer: A report from the childhood cancer survivor study. J Natl Cancer Inst 2006;98(13):890-896.

11. Green D, Sklar C, Boice J, et al. Ovarian failure and reproductive outcomes after childhood cancer treatment: Results from the childhood cancer survivor study. J Clin Oncol 2009;27(14):2374-2381.

12. Thibaud E, Ramirez M, Brauner R, et al. Preservation of ovarian function by ovarian transposition performed before pelvic irradiation during childhood. J Pediatr 1992;121(6):880-884.

13. Green D,Nolan V,Goodman P,et al. The cyclophosphamide equivalent dose as an approach for quantifying alkylating agent exposure: A report from the childhood cancer survivor study. Pediatric blood & cancer 2014;61(1):53-67.

14. te Velde E, Pearson P. The variability of female reproductive ageing. Hum Reprod Update 2002;8(2):141-154.

15. Metzger M, Meacham L, Patterson B, et al. Female reproductive health after childhood, adolescent, and young adult cancers: Guidelines for the assessment and management of female reproductive complications. J Clin Oncol 2013;31(9):1239-1247.

16. Breiman L. Random forests. Mach Learning 2001;45(1):5-32.

17. Sowa J, Heider D, Bechmann L, et al. Novel algorithm for non-invasive assessment of fibrosis in NAFLD. PLoS One 2013;8(4):e62439.

18. Taylor R, Pare J, Venkatesh A, et al. Prediction of in hospital mortality in emergency department patients with sepsis: A local big data driven, machine learning approach. Acad Emerg Med (In press).

19. Schöfl G, Schmidt A, Lange V. Prediction of spurious HLA class II typing results using probabilistic classification. Hum Immunol (In press).

20. Therneau T, Grambsch P. Modeling survival data: Extending the Cox model. 1st ed. Springer; 2000.

21. Lawless J. Statistical models and methods for lifetime data. 2nd ed. Wiley; 2003.

22. Zhou Q, Zheng Y, Chibnik L, et al. Assessing incremental value of biomarkers with multiphase nested case control studies. Biometrics 2015;71(4):1139-1149.

23. Hothorn T, Lausen B, Benner A, et al. Bagging survival trees. Stat Med 2004;23(1):77-91.

24. Rotnitzky A, Robins J. Inverse probability weighting in survival analysis. 2nd ed. The Encyclopedia of Biostatistics. Wiley; 2005.

25. Zhang H, Singer B. Recursive partitioning in the health sciences. 1st ed. Springer; 2013.

26.Tibshirani R. Regression shrinkage and selection via the lasso. JRSSB 1996:267-288.

27. Yuan Y, Su W, Zhu M. Threshold-free measures for assessing the performance of medical screening tests. Front Public Health 2015;3:57.

28. 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(8):855-859.

29. Pepe M. The statistical evaluation of medical tests for classification and prediction. Oxford University Press; 2003.

30. Steyerberg E, Vickers A, Cook N, et al. Assessing the performance of prediction models: A framework for traditional and novel measures. Epidemiology 2010;21(1):128-138.

31. Gneiting T, Balabdaoui F, Raftery A. Probabilistic forecasts, calibration and sharpness. JRSSB 2007;69(2):243-268.

32. Yuan Y, Zhou Q, Li B, Cai H, Chow E, Armstrong G. (2016) A threshold-free prospective prediction accuracy measure for censored time to event data. http://arxiv.org/abs/1606.04172. arXiv:1606.04172v2 [stat.ME]

33. Heagerty P, Zheng Y. Survival model predictive accuracy and ROC curves. Biometrics 2005;61(1):92-105.

Appendix

Letter of Support from St. Jude Lifetime Cohort study (SJLIFE)

St. Jude Children's Research Hospital
ALSAC · Danny Thomas, Founder

February 26, 2016

Yan Yuan, Ph.D.
Assistant Professor
School of Public Health, University of Alberta
3-299 ECHA
Edmonton, Alberta CANADA T6G 1C9

Dear Dr. Yuan,

I am writing as the Principal Investigator of the St. Jude Lifetime Cohort study to offer my enthusiastic support of your proposed research project entitled, "Risk prediction for ovarian failure in childhood cancer survivors". I am pleased to collaborate on this important project and to provide you with access to our cohort study data for validation.

The St. Jude Lifetime Cohort consists of more than 3000 survivors who are regularly followed with onsite clinical assessments at St. Jude Children's Research Hospital. The unique strength of our study/database is the availability of detailed cancer treatment information and up-to-date, comprehensive clinically-assessed late health outcomes for all study participants. The in-kind support is valued at approximately USD1,000 for staff time on pulling the dataset, creating a data dictionary specifically for your project and corresponding with your team. I understand that this data will serve as an external validation cohort for the prediction tools to be developed by your proposed research.

By providing a tool that will stratify the risk of infertility, I believe that your proposed work will add great value to clinical practice and enhance survivors' quality of life by facilitating appropriate fertility consultations as they enter adulthood. I look forward to this collaboration.

Sincerely,

Melissa M. Hudson, M.D., Director
Division of Cancer Survivorship
Department of Oncology
St. Jude Children's Research Hospital

262 Danny Thomas Place, Memphis, TN 38105-3678 | t 901.595.3300 | www.stjude.org