Predicting age-specific primary ovarian insufficiency risk: A report from the Childhood Cancer Survivor Study (CCSS) and St. Jude Lifetime Cohort Study (SJLIFE)

Zhe Lu (MSc)[1], Jessica Baedke (MSc)[2], Lin Yu (MSc)[1], Sogol Mostoufi-Moab (MD, MSCE)[3], Angela Delany (MD)[2], Paul C Nathan (MD, MSc, FRCPC)[4], Yutaka Yasui (PhD)[2], Eric J Chow (MD, MPH)[5], Rebecca M Howell (PhD)[6], Gregory T Armstrong (MD, MSCE)[2], Smita Bhatia (MD, MPH)[7], Cindy Im (PhD)[1], Yan Yuan (PhD)[1]
1 University of Alberta, Edmonton, AB Canada
2 St. Jude Children's Research Hospital for Children, Memphis, TN USA
3 The Children's Hospital of Philadelphia, Philadelphia, PA USA
4 The Hospital for Sick Children, Toronto, ON Canada
5 Seattle Children's Hospital, Seattle, WA USA
6 MD Anderson Cancer Center, Houston, TX USA
7 University of Alabama at Birmingham, Birmingham AL USA

Current word count: 291 (Limit 300)

Aim: To develop and validate prediction algorithms for age-specific risk of experiencing primary ovarian insufficiency (POI), defined as natural menopause before age 40, in female survivors of childhood cancer.

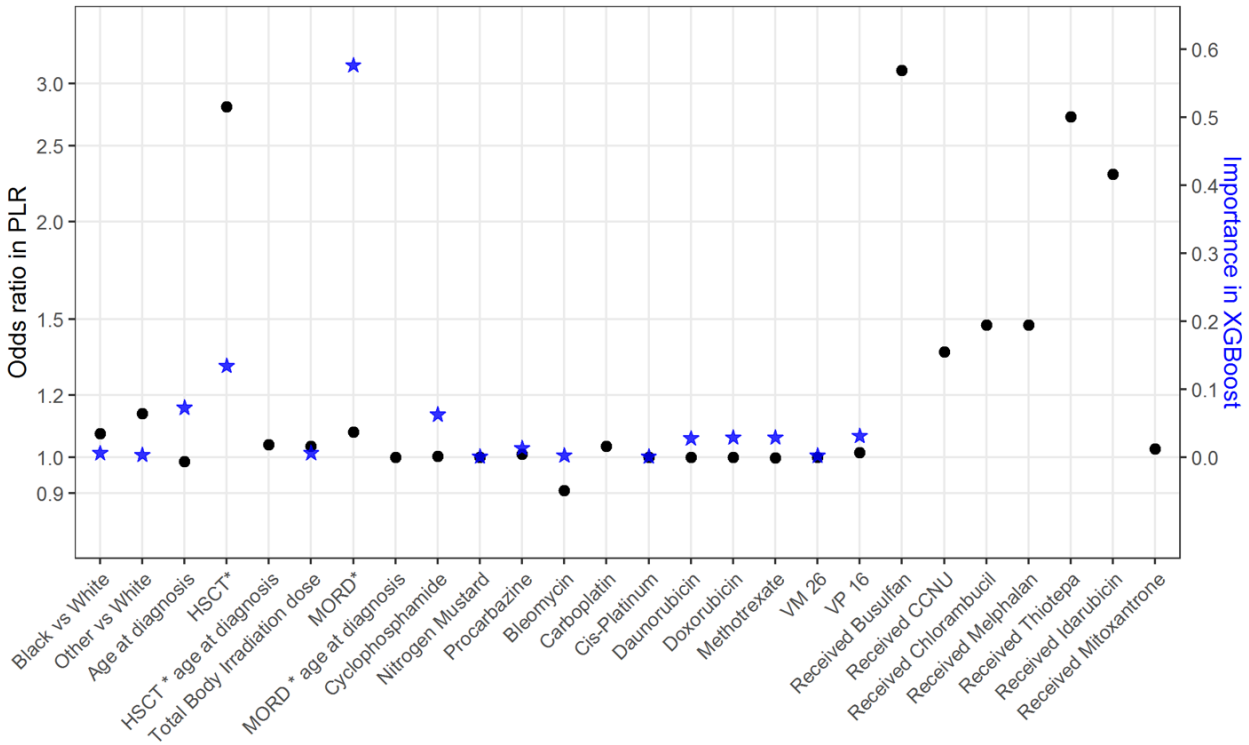
Background and methods: Survivors are at increased risk of POI. To counsel survivors regarding their individualized needs for fertility preservation, accurate risk estimates at different ages are needed. Data from 7891 (CCSS) survivors were used to develop prediction algorithms and data 1351 (SJLIFE) survivors were used for external validation. Ovarian status was ascertained from longitudinal self-reported menstrual surveys (CCSS) or clinical assessment (SJLIFE). We used penalized logistic regression and XGBoost to develop separate POI risk algorithms as survivors aged from 21 to 40 years. For a subset of 1985 CCSS participants with genotype data, we evaluated the contributions of polygenic risk scores (PRSs) from published general population genome-wide association studies of natural menopause age to predict POI risk by age 40.

Results: Both penalized logistic regression and XGBoost performed well, with the XGBoost-based algorithms performed better. By age 30, the most influential predictors are minimum ovarian radiation dose, haematopoietic stem-cell transplant, diagnosis age, and cyclophosphamide dose (Figure 1). As survivors aged from 21 to 40 years, the POI prevalence increased from 7.9% to 18.6% (CCSS) and 8.8% to 16.8% (SJLIFE). Across this age range, algorithm AUCs were 0.76-0.80 in the development sample (CCSS) and 0.85-0.90 in the validation sample (SJLIFE), while the average precision increased from 0.45 to 0.60 (CCSS) and 0.58 to 0.80 (SJLIFE). Adding PRSs improved

calibration (Spiegelhalter-z decreased from 11.4 to 0.1), but not the prediction performance.

Conclusions: Given cancer diagnosis and treatment information, our POI risk prediction algorithms can inform decisions regarding the need for fertility preservation interventions among female childhood cancer survivors. An APP has been developed to visualize POI risk as survivors age.

Figure 1: Estimated odds ratio from penalized logistic regression (PLR) and predictors' importance assessed by XGBoost for predicting cancer survivors' POI risk by age 30.



*HSCT: Haematopoietic stem-cell transplant; MORD: Minimum ovarian radiation dose
The units for radiation dose and chemotherapy agent doe are Gy and g/m^2, respectively.
Doses of Procarbazine and Nitrogen Mustard were converted to Cyclophosphamide Equivalent Dose (BCNU and Ifosfamide were not listed in the x-axis as they were not selected by PLR or XGBoost). Rarely used Busulfan, CCNU, Chlorambucil, Melphalan, Thiotepa, Idarubicin, and Mitoxantrone were converted to binary variables (received or not) in PLR.