Study Title: International Study of Subsequent Colorectal Cancer Among Survivors of Childhood, Adolescent and Young Adult Cancers (I-SCRY)

MPI: Chaya Moskowitz / Tara Henderson

Working Groups: SMN (Primary); Cancer Control (Secondary)

Background and Rationale:
Treatment advances in pediatric cancer have led to a growing population of long-term survivors of childhood, adolescent, and young adult (AYA) cancer that exceeds 500,000 in the U.S. These survivors have an elevated risk of subsequent malignant neoplasms (SMN), leading to premature mortality in this population. One of the SMNs in this population that has not received significant attention is colorectal cancer (CRC). Yet, several recent studies suggest that there is a substantially increased risk of CRC in childhood and AYA cancer survivors. Understanding which childhood and AYA cancer survivors are at highest risk for developing CRC is critical in order to ensure they are screened with colonoscopy, which can both detect and prevent this SMN, and ultimately minimize its associated morbidity and premature mortality.

The Children's Oncology Group recommends screening beginning at age 30 or 5 years after treatment, whichever occurs later, in childhood cancer survivors treated with any dose of abdominal or pelvic RT and irrespective of other risk factors or treatment exposures. Risk of CRC appears especially high for survivors treated with abdominal or pelvic radiotherapy (RT) with standardized incidence ratios (SIR) suggesting that CRC risk is potentially 2- to 8-fold higher relative to the general population, although risk may vary by age at radiation exposure. Importantly, our knowledge of how other treatment exposures (including detailed radiation and chemotherapy exposures) may impact CRC risk is very limited. In particular, it has been hypothesized that treatment with cisplatin or procarbazine, and the interaction between these agents and abdominal/pelvic RT, may contribute to increased CRC risk, but previous studies have been unable to comprehensively address these associations. Similarly, it is unknown whether the prognosis, particularly mortality, after a treatment-related CRC differs from mortality following a *de novo* CRC diagnosis; this information is vital to thoroughly evaluate screening strategies (e.g., cost-effectiveness) in this population in the future. These critical gaps in our knowledge, which are barriers to evidenced-based prevention of CRC, are due to the limited number of CRCs in any single study cohort, the lack of detailed information on primary cancer treatments for some cohorts, and the relatively young age of the cohorts.

Thus, we propose to leverage the unique resources of eight international childhood and AYA cancer survivor cohorts. We will pool individual participant data on 51,309 survivors, at least 298 with CRCs, with median follow-up across studies ranging from 16 to 26 years after primary cancer diagnosis, to comprehensively examine risk factors of, and mortality after, subsequent CRCs. Of note, this is more than 3-fold the number of CRCs in any known childhood or AYA cohort with detailed treatment exposure data available. Thus, the combination of childhood and AYA cancer survivors into a single cohort is novel and given common treatment approaches to pediatric and AYA cancers, will enable us to address these questions with the ability to focus on the role of age at exposure. The combined data will be used to address the following specific aims:

Primary Aim. To evaluate the association of childhood and AYA cancer therapy (abdominal and/or pelvic radiation fields and dose, chemotherapy exposures including cumulative doses of alkylating agents, anthracyclines, platinating agents) on CRC risk in a large, multi-cohort, international survivor population.

Secondary Aims. To conduct an expanded assessment of the burden of CRC in childhood and AYA cancer survivors and compare it to the general population in a large, multi-cohort, international survivor population:

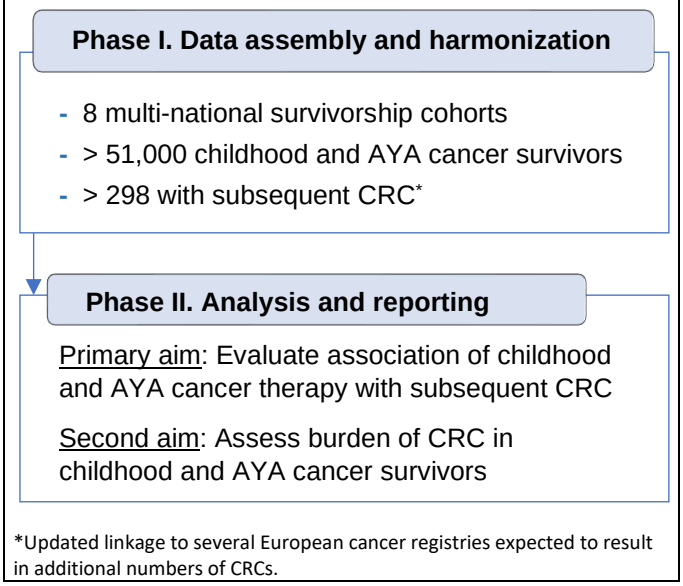
i. Using SIRs and absolute excess risk, generate precise estimates of the incidence of CRC in childhood and AYA cancer survivors relative to the age- and sex-matched general population of the country of origin overall and by primary cancer diagnosis, sex, age at primary cancer diagnosis, attained age, and time since primary cancer diagnosis.
ii. Estimate the cumulative incidence of subsequent CRC by primary cancer diagnosis and by age at primary cancer diagnosis.
iii. Determine whether childhood and AYA cancer survivors with subsequent CRC have higher overall and cause-specific mortality compared to mortality following *de novo* CRC in the general population.

This study will address important clinical questions including providing a basis for refining surveillance guidelines and enabling strong cost-effectiveness analyses of these guidelines.

Study Overview

This study is a collaborative effort between multiple study groups representing survivors of a childhood and AYA cancer in North America and Europe. We propose to conduct an individual participant data (IPD) meta-analysis and pool data on individual participants in 8 existing cohorts. These cohorts were selected for inclusion based on primary cancer diagnoses defining the cohort, ages at which participants were diagnosed with these primary cancers, availability of high quality, detailed treatment information for the primary cancer, and numbers of CRC cases observed within the cohort. In the first phase of this study, data on approximately 51,000 individuals diagnosed with a cancer prior to age 40 will be sent to Memorial Sloan Kettering Cancer Center (MSKCC). The data items will be standardized to common scales of measurements across studies to permit unified analyses addressing our hypotheses. The second phase of this study will involve analyzing the pooled data to address the scientific questions of interest. The overall goal of the work proposed in this application is to elucidate the associations between subsequent CRC and treatments used for a childhood and AYA cancer, particularly to better understand how the interaction of these treatments may modify the risk of subsequent CRC, to understand the role that age at treatment plays in the risk of CRC, and to examine prognosis following subsequent CRC in this population. Across all the cohorts, key components of the RT treatment were abstracted from MRs and RT records in a similar manner. Specifically, information on the prescriptions, including the treatment fields and delivered doses, as well as the dates that RT was administered, were recorded. What we will not have available is dosimetry estimates of the absorbed dose of RT at the CRC tumor site. Dosimetry is not available in clinical care, though. The delivered doses and fields will be more useful for informing screening guidelines.

Figure 2. Study Overview



C2.3 Description of the available data

In preparation for this proposal, we queried the individual cohort databases to ascertain the numbers of survivors with CRC and key treatment exposures. As shown in **Table 1**, there are a total of 51,309 childhood and AYA cancer survivors who will contribute to the pooled cohort. Across the cohorts there are currently 298 survivors who were subsequently diagnosed with CRC. Note that prior to sending data to MSKCC, several of the European study groups (the Dutch and Norwegian groups) will update data and perform new linkages to their cancer registries, so the number of CRCs is expected to increase. Median follow-up after the primary cancer diagnosis across the cohorts ranges from 13 to 26 years. Over 43,000 were alive at last contact and 134 have died after being diagnosed with CRC. 42% are female. With regard to ages at primary cancer diagnosis: 22,183 were diagnosed before age 10 (62 with CRC), 13,092 were diagnosed between ages 10 and 19 (65 with CRC), and 16,034 were diagnosed at age 20 or older (171 with CRC).

We have also prepared preliminary counts on the number of cohort members with key treatment exposures of interest. As shown in **Table 2**, 20% (n=10,374) of the pooled cohort were treated with abdominal or pelvic RT, 12% (n=5,987) with procarbazine and 11% (n=5,575) with cisplatin.

Finally, we also queried the individual cohort databases for information on smoking history and body mass index (BMI). Because this is information that changes over time and is not coded consistently across the cohorts, our ability to summarize it thoroughly before having data and harmonizing it is limited. Based on our queries for information on when survivors were between the ages of 25 and 30 or as near as possible to that, we estimate that the percentages of survivors who reported ever smoking range from about 23% (the U.S. cohorts) to approximately 40% (European cohorts). Furthermore, between 30% and 50% of the cohorts' participants were overweight or obese (BMI ≥ 25) during this same time frame. Dietary and alcohol consumption is not available consistently in the cohorts.

Table 1. Description of cohorts

Cohort	Age range at Dx, years	Years of Dx	Years follow-up, median (IQR)	Frequency		Vital status		
				Total	CRC	Alive	Deceased without CRC	Deceased after CRC
CCSS*	0-21	1970-1999	25 (19, 33)	22,265	80	18,648	3,591	26
SJLIFE	0-21	1962-2012	24 (16, 33)	5,017	17	4,801	210	6
LATER	0-17	1963-2001	25 (19, 33)	5,843	13	5,359	482	2
Dutch HL cohort	0-40	1965-2008	21 (15, 29)	2,910	37	1,916	966	28
CRYSTAL	15-40	1989-2012	13 (8, 18)	790	5	**	**	**
Dutch testicular ca.	10-40	1976-2007	16 (11, 22)	4,635	26	4,373	253	9
UK HL cohort	0-36	1956-2003	26 (20, 34)	4,380	55	3,246	1,100	36
Norwegian testicular ca.	27-40	1980-2009	19 (13,26)	5,469	65	4,849	591	27
Total				**51,309**	**298**	**43,192**	**7,193**	**134**

Dx = Diagnosis; IQR = interquartile range; ca.= cancer cohort; *Survivors who participate in both CCSS and SJLIFE have been removed from the tabulations for CCSS. **The Dutch NHL cohort is being cleaned as this application is being prepared and this information is not yet available.

Table 2. Frequencies of treatment exposures, overall and for patients with colorectal cancer, across cohorts

Cohort	Abdominal/ Pelvic RT		Alkylating agents		Procarbazine		Cisplatin		Anthracyclines	
	Total	CRC	Total	CRC	Total	CRC	Total	CRC	Total	CRC
CCSS	4,519	36	10,831	58	1,863	28	1,928	7	9,374	35
SJLIFE	1,132	8	2,827	12	241	2	456	2	2,827	10
LATER	1,006	7	2,987	4	407	2	433	0	2,690	2
Dutch HL cohort	1,188	27	1,911	22	1,470	20	N/A*	N/A*	1,387	3
CRYSTAL	81	0	714	4	N/A*	N/A*	59	0	737	5
Dutch testicular ca.	332	14	0	0	N/A*	N/A*	408	9	0	0
UK HL cohort	472	12	2,302	38	2,006	35	61	0	1,580	11
Norwegian testicular ca.	1,644	37	41	0	0	0	2,230	21	<10	0
Total	**10,374**	**141**	**21,613**	**138**	**5,987**	**67**	**5,575**	**39**	**18,595**	**66**

RT = radiotherapy; ca = cancer; *Treatment very rarely used in this cohort.

Data to be requested

We plan for de-identified datasets to be sent to MSKCC via a secure password-protected file transfer protocol service. Raw data will be stored in its original format. The list of data elements to be obtained from each study group is shown in **Table 3**.

Table 3. Data elements to be assembled across the cohorts

Primary cancer diagnosis (diagnosis and date of diagnosis)
Age at primary cancer diagnosis
Sex
Race
Radiotherapy (field, delivered dose, and date started)
Anthracyclines (specific drugs, cumulative doses, and date started)
Alkylating agents (specific drugs, cumulative doses, and date started)
Platinum-based drugs (specific drugs, cumulative doses, and date started)
Anti-metabolites (specific drugs, cumulative doses, and date started)
Anti-tumor antibiotics (specific drugs, cumulative doses, and date started)
Corticosteroids (specific drugs, cumulative doses, and date started)
Plant alkaloids (specific drugs, cumulative doses, and date started)
Epipodophyllotoxins (specific drugs, cumulative doses, and date started)
Enzyme chemotherapy (specific drugs, cumulative doses, and date started)
Colorectal cancer diagnosis (including stage and location)
Date and age at colorectal cancer diagnosis
Vital status (alive/dead, date of death or last contact, and cause of death)
Smoking status and age at which the information was captured
Body mass index and age at which the information was captured

General statistical considerations

This study is a retrospective individual participant data (IPD) meta-analysis that combines hospital- and population-based data. Before describing the analytic methods to be used for the individual Aims, here we describe methodology applicable to all Aims. Analyses will begin by using descriptive methods such as summary statistics (means, medians, etc.) and graphical methods to explore and understand trends. This will be done for the entire pooled cohort and separately by individual cohorts to understand between-study heterogeneity. Our plan is to use a one-stage approach, analyzing individual-level data from the different cohorts in a single step while accounting for clustering of individuals within studies. This approach allows all studies to be analyzed simultaneously, provides

greater flexibility in modeling choices, greater ability to study interactions, and can have better convergence properties if some studies contribute a small number of cases.[1, 2] For all analyses, childhood and AYA cancer survivors will be considered at risk of CRC beginning at five years after their primary cancer diagnosis until death, a CRC diagnosis, or date of last contact. In general, the time scale for analysis will be the time since the primary cancer diagnosis, but we will also present key results using age as the time scale.

Heterogeneity of effects between studies: We will carefully examine between-study heterogeneity and participant-level variation in associations between treatment exposures and risk of CRC. Possible sources of heterogeneity include different incidence rates of CRC in the different countries, different CRC surveillance patterns since recommendations for this population vary across countries, different follow-up times between the studies, and different distributions of effect modifiers across studies. In the methods detailed below, we plan to use mixed effects (hierarchical) models to account for between-study heterogeneity and clustering of patients within study. Within this framework, we will evaluate causes of heterogeneity. For potential causes of heterogeneity that are constant within study (e.g. country or screening recommendation), mixed effects models will be used to test the null hypothesis of no effect-modification by including an interaction between potential effect modifiers and treatment exposures. Covariates will be centered at within-study mean values to avoid ecological bias.[1, 3] Similarly, mixed effects models will be used to evaluate potential causes of heterogeneity that vary within study (e.g. smoking status) by including an interaction term and including random effects to allow for residual heterogeneity.[3] In the case of different lengths of follow-up between studies and effects that change over time, potential heterogeneity will be explored by modeling the effect as time-dependent.[1, 4]

Primary Aim: Evaluate the association of childhood and AYA cancer therapy with CRC

We will evaluate the association between CRC risk and childhood and AYA cancer therapy. We hypothesize that abdominal/pelvic RT, procarbazine, and cisplatin will be associated with elevated CRC risks and that there is dose-response relationship between doses and risk. We hypothesize there is an interaction between chemotherapies and abdominal/pelvic RT such that the association between procarbazine or cisplatin and CRC risk is different depending upon RT exposure. We further hypothesize that age at exposure modifies the risk conferred by these therapies.

The models will use the time since diagnosis (or more specifically, the time from diagnosis + 5 years) as the time scale and adjust for other covariates including age at treatment, sex, race, smoking status, and BMI. We will use mixed effects, stratified, cause-specific, Cox proportional hazards models to evaluate associations.[3] This approach will allow us to simultaneously estimate heterogeneity of the baseline rate of CRC within studies and heterogeneity of the treatment exposure effects. We will specify a separate baseline hazard for each study, allowing the underlying baseline risk of CRC to differ by study (without assuming proportional hazards).

Secondary Aims: Assess the burden of CRC in childhood and AYA cancer survivors

To compare the incidence of CRC in childhood and AYA cancer survivors to that observed in the general population, SIRs, defined as the ratio of the observed incidence rates of childhood and AYA cancer survivors with CRC relative to the incidence rate of CRC in the general population, will be estimated. Age-, calendar year-, and sex-specific incidence rates in the general population will be obtained separately for each cohort using registries in the respective countries. For CCSS and SJLIFE, population incidence rates of CRC in the US will be obtained from the Surveillance, Epidemiology and End Results (SEER) program.[5] For the Dutch cohorts, population incidence rates will be obtained from the Netherland Cancer Registry[6], for the UK cohort, from the Office of National Statistics[7], and for the Norwegian cohort from the Cancer Registry of Norway[8]. These analyses will be carried out using a multilevel Poisson regression framework with a log link function that includes an offset term for the log of the expected counts as well as study-specific random effects to account for within-study clustering.[9]

We will also estimate the absolute risk of CRC using mixed effects models. The cumulative incidence of CRC treating death without CRC as a competing risk will be estimated using an estimator of cumulative incidence arising from a random effects proportional hazards model that accounts for the within-study clustering. This is a model for the subdistribution hazard (similar to the frequently used Fine and Gray approach for competing risks regression), but incorporates a frailty term (random effect) representing the different studies.[10]

Overall mortality after a subsequent CRC diagnosis in childhood and AYA cancer survivors will be analyzed using the mixed-effects stratified Cox model described above for Aim 2. Participants will be considered at risk of death starting at the date of diagnosis with CRC. Time since the CRC diagnosis will be the time scale.

References:

1. de Jong VMT, Moons KGM, Riley RD, Tudur Smith C, Marson AG, Eijkemans MJC, Debray TPA. Individual participant data meta-analysis of intervention studies with time-to-event outcomes: A review of the methodology and an applied example. Res Synth Methods. 2020;11(2):148-68. Epub 2020/02/06. doi: 10.1002/jrsm.1384. PubMed PMID: 31759339.

2. Burke DL, Ensor J, Riley RD. Meta-analysis using individual participant data: one-stage and two-stage approaches, and why they may differ. Stat Med. 2017;36(5):855-75. Epub 2016/10/16. doi: 10.1002/sim.7141. PubMed PMID: 27747915.

3. Hua H, Burke DL, Crowther MJ, Ensor J, Tudur Smith C, Riley RD. One-stage individual participant data meta-analysis models: estimation of treatment-covariate interactions must avoid ecological bias by separating out within-trial and across-trial information. Stat Med. 2017;36(5):772-89. Epub 2016/12/03. doi: 10.1002/sim.7171. PubMed PMID: 27910122; PMCID: PMC5299543.

4. Crowther MJ, Riley RD, Staessen JA, Wang J, Gueyffier F, Lambert PC. Individual patient data meta-analysis of survival data using Poisson regression models. BMC Med Res Methodol. 2012;12:34. Epub 2012/03/27. doi: 10.1186/1471-2288-12-34. PubMed PMID: 22443286; PMCID: PMC3398853.

5. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov). SEER*Stat Database: Incidence - SEER Research Data, 9 Registries, Nov 2019 Sub (1975-2017) - Linked To County Attributes - Time Dependent (1990-2017) Income/Rurality, 1969-2017 Counties, National Cancer Institute, DCCPS, Surveillance Research Program. Released April 2020, based on the November 2019 submission.

6. Bray F CM, Mery L, Piñeros M, Znaor A, Zanetti R and Ferlay J, editors. Cancer Incidence in Five Continents. Lyon: International Agency for Research on Cancer; 2017 Available from: https://ci5.iarc.fr. .

7. Office of National Statistics. Cancer Statistics Registrations - Series MB1. The Stationary Office. London, England2006.

8. Cancer Registry of Norway. Cancer In Norway 2016 - Cancer incidence, mortality, survival, and prevalence in Norway. Oslo: Cancer Registry of Norway; 2017.

9. Austin PC, Stryhn H, Leckie G, Merlo J. Measures of clustering and heterogeneity in multilevel Poisson regression analyses of rates/count data. Stat Med. 2018;37(4):572-89. Epub 2017/11/08. doi: 10.1002/sim.7532. PubMed PMID: 29114926.

10. Katsahian S, Resche-Rigon M, Chevret S, Porcher R. Analysing multicentre competing risks data with a mixed proportional hazards model for the subdistribution. Stat Med. 2006;25(24):4267-78. Epub 2006/09/09. doi: 10.1002/sim.2684. PubMed PMID: 16960919.