Integration of Cancer Registry and Electronic Health Record Data to Construct a Childhood Cancer Survivorship Cohort, Facilitate Risk Stratification, and Assess Appropriate Follow-up Care

David Noyd[1], Nigel Neely[2], kristin Schroeder[3], Paul Lantos[1], Steve Power[2], Susan Kreissman[3], and Kevin Oeffinger[2]

[1]Duke University Medical Center
[2]Duke Cancer Institute
[3]Duke University Hospital

October 29, 2020

Abstract

Background: This retrospective study harnessed an institutional cancer registry to construct a childhood cancer survivorship cohort, integrate electronic health record (EHR) and geospatial data to risk stratify patients for serious adverse health outcomes, analyze follow-up care patterns, and determine factors associated with suboptimal follow-up care. Procedure: The survivorship cohort included patients [?]18 years of age with a diagnosis of a malignancy reported to the institutional cancer registry between January 1, 1994 and November 30, 2012. ICD-O-3 coding and treatment exposures facilitated risk stratification of survivors. All follow-up visits were extracted from the EHR through linkage to the cancer registry based on medical record number (MRN). Results: Eight-hundred-and-sixty-five survivors were included in the final analytic cohort, of whom 191, 496, and 158 were considered low, intermediate, and high risk survivors, respectively. Two-hundred-and-eight-two survivors (32.6%) were not seen in any oncology-related subspecialty clinic at Duke five to seven years after initial diagnosis. Factors associated with a clinic visit included younger age (p=0.008), acute lymphoblastic leukemia (ALL) as the primary diagnosis (p<0.001), race/ethnicity (p=0.010), risk strata (p=0.001), distance to treatment center (p<0.0001), and lower ADI (p=0.011). Multivariable logistic modeling with adjustment for diagnosis of ALL, gender, age at diagnosis, and race/ethnicity attenuated the association between follow-up care and risk strata (p=0.17) Conclusions: Nearly a third of survivors received suboptimal follow-up care. This study provides a reproducible model to integrate cancer registry and EHR data to construct risk-stratified survivorship cohorts to assess follow-up care.

Integration of Cancer Registry and Electronic Health Record Data to Construct a Childhood Cancer Survivorship Cohort, Facilitate Risk Stratification, and Assess Appropriate Follow-up Care

David H. Noyd, MD, MPH[1*]; Nigel B. Neely, MSc[2]; Kristin M. Schroeder, MD, MPH[1]; Paul M. Lantos, MD, MS[3]; Steve Power, MBA[2]; Susan G. Kreissman, MD[1]; Kevin C. Oeffinger, MD[2,4]

[1] Department of Pediatrics, Division of Pediatric Hematology-Oncology, Duke University Medical Center

[2]Duke Cancer Institute

[3]Departments of Internal Medicine, Pediatrics, and Community/Family Medicine, Duke University Medical Center

[4]Department of Medicine, Duke University Medical Center

*Correspondence:

David H. Noyd, MD, MPH

DUMC Box 102382, 382 Hanes House

330 Trent Drive, Durham, NC 27710, USA

Telephone +1 919 681 1624

Fax +1 919 681 7950

Email: david.noyd@duke.edu

Abstract Word Count: 240 (Max 250)

Main Text Word Count: 3499 (Max 3500)

Number of Tables, Figures, and Legends: 5 (Max 6)

Short Running Title: Childhood Cancer Survivor Risk Stratification and Follow-up

Key Words: Childhood Cancer Survivorship, Biomedical informatics

Abbreviations Table:

ADI	Area Deprivation Index	ICD-O-3	International Classification of Diseases for Oncology, th
EHR	Electronic Health Record	ICCC-3	International Classification of Childhood Cancer, third
ALL	Acute Lymphoblastic Leukemia	PHO	Pediatric Hematology-Oncology
MRN	Medical Record Number	PNO	Pediatric Neuro-Oncology
NCDB	National Cancer Database	PBMT	Pediatric Blood and Marrow Transplant
RUCA	Rural-Urban Commuting Area	FORDS	Facility Oncology Registry Data Standards
NC,SC,VA	North Carolina, South Carolina, Virginia		

ABSTRACT

Background: This retrospective study harnessed an institutional cancer registry to construct a childhood cancer survivorship cohort, integrate electronic health record (EHR) and geospatial data to risk stratify patients for serious adverse health outcomes, analyze follow-up care patterns, and determine factors associated with suboptimal follow-up care.

Procedure: The survivorship cohort included patients [?]18 years of age with a diagnosis of a malignancy reported to the institutional cancer registry between January 1, 1994 and November 30, 2012. ICD-O-3 coding and treatment exposures facilitated risk stratification of survivors. All follow-up visits were extracted from the EHR through linkage to the cancer registry based on medical record number (MRN).

Results: Eight-hundred-and-sixty-five survivors were included in the final analytic cohort, of whom 191, 496, and 158 were considered low, intermediate, and high risk survivors, respectively. Two-hundred-and-eight-two survivors (32.6%) were not seen in any oncology-related subspecialty clinic at Duke five to seven years after initial diagnosis. Factors associated with a clinic visit included younger age (p=0.008), acute lymphoblastic leukemia (ALL) as the primary diagnosis (p<0.001), race/ethnicity (p=0.010), risk strata (p=0.001), distance to treatment center (p<0.0001), and lower ADI (p=0.011). Multivariable logistic modeling with adjustment for diagnosis of ALL, gender, age at diagnosis, and race/ethnicity attenuated the association between follow-up care and risk strata (p=0.17)

Conclusions: Nearly a third of survivors received suboptimal follow-up care. This study provides a reproducible model to integrate cancer registry and EHR data to construct risk-stratified survivorship cohorts to assess follow-up care.

1 | INTRODUCTION

Remarkable progress in the overall survival of children with cancer over the past several decades emphasizes the importance of long-term follow-up care for survivors.[1] Indeed, the late effects of treatment, from secondary malignancies to cardiovascular disease, affect every major organ system and demonstrate significant heterogeneity in incidence based on underlying malignancy and specific exposures.[2] Clinical guidelines offer a framework for disease surveillance after cancer treatment, yet lack of follow-up is a significant barrier to early detection of late effects. The electronic health record (EHR) sparks unique approaches to construct, maintain, and leverage childhood cancer survivorship cohorts to optimize health care delivery and as a platform for survivorship research. Ascertainment of treatment exposures aids in risk stratification and identification of survivors lost to follow-up.

Risk stratification of patients to determine appropriate levels of follow-up care, both for the coordination with primary care providers as well as the identification of high-risk survivors who would benefit from regular subspecialist visits, facilitates judicious use of health care resources to optimize patient outcomes.[3] In the United Kingdom, different levels of follow-up care were proposed based on treatment exposures[4] were later validated to ensure appropriate capture of adverse effects.[5] The British Childhood Cancer Survivor Study (BCCSS) further refined the risk stratification through the integration of underlying malignancy.[6] The introduction of novel targeted agents and immunotherapy underscores the need for an adaptable system to incorporate new risks for late effects from treatment. Albeit an emerging field, there are associated cardiac,[7]renal,[8] and endocrinologic[9] adverse effects during and shortly after treatment with these therapeutic agents.[10] In the coming years, new late effects from these modalities will likely impact the care of adult survivors of childhood cancer.

As knowledge of late effects risk improves, guideline recommendations offer a framework for disease surveillance in order to mitigate further complications. The Children's Oncology Group (COG) regularly updates long-term follow-up guidelines based on treatment exposure, including radiation, surgery, immunotherapy, and each specific chemotherapeutic agent.[11] Furthermore, international efforts to harmonize different recommendations across the globe provide another layer for evidence-based survivorship care.[12] Over the past decade, published recommendations for cardiomyopathy,[13] breast cancer,[14]fertility,[15,16], thyroid cancer,[17]and ototoxicity[18] address major challenges for survivors and inform delivery of appropriate care for early detection of these late effects. Large, multi-institutional longitudinal studies, such as the Childhood Cancer Survivorship Study (CCSS),[19]offer a wealth of data to inform validated risk prediction models for acute ovarian failure,[20]thyroid cancer,[21] and cardiovascular outcomes.[22,23] Translation and implementation of these recommendations and models into clinical practice, on an individual patient and population health level, represent a tremendous opportunity to optimize health outcomes in the digital age.

Long-term follow-up care of survivors enhances patient education and early detection of late effects.[24] Predictors of suboptimal follow-up, such as age, insurance status, and race/ethnicity, help identity at-risk populations.[25] Studies of adult cancer patients from rural areas reveal significant disparities with regards to overall mortality[26-28] and financial toxicity[29]; however, there is a paucity of data for survivors of childhood cancer. During active treatment, distance to the cancer center is a considerable barrier,[30,31]especially for adolescents and young adults.[32] For high-risk survivors, distance to a comprehensive cancer center may also influence their likelihood of receiving appropriate follow-up care.

The primary aim of this study was to harness an institutional cancer registry, which follows the standards of and reports to the National Cancer Database (NCDB),[33] to construct a childhood cancer survivorship cohort, integrate EHR and geospatial data to analyze follow-up care patterns, and determine factors associated with inadequate follow-up care. Risk stratification of survivors provides an additional lens to prioritize survivors at risk for inadequate follow-up care.

2 | METHODS

| Cancer registry and patient information

Construction of this single institution childhood cancer survivorship cohort was based on cancer registry data and integration of EHR data elements linked by the medical record number (MRN). As part of the accreditation by the American College of Surgeons Commission on Cancer, centers are required to report all newly diagnosed cases to the NCDB.[33] Centers are also required to report cases to respective state registries regardless of their accreditation status. Inclusion criteria for the construction of this childhood cancer survivorship cohort (Figure 1) included patients [?]18 years of age with a diagnosis of a malignancy reported to the cancer registry. The cohort was limited to patients seen in the pediatric oncology (PHO) or neuro-oncology (PNO) clinics between January 1, 1994 and November 30, 2012 in order to ensure a seven year follow-up period for all survivors. Patients who died or had a documented relapse during the seven year follow-up window after date of diagnosis were excluded. In order to exclude referrals for refractory or relapsed cases that were not diagnosed and treated at this institution, only analytic cases were included. Analytic cases are defined by the Facility Oncology Registry Data Standards (FORDS) Manual as cases diagnosed at and/or received all or part of the first course of treatment at the reporting facility (Duke University Medical Center).

| Disease classification

Cancer diagnoses were grouped according to the International Classification of Childhood Cancer, third revision (ICCC-3),[34] by using the International Classification of Diseases for Oncology, third revision (ICD-O-3), as reported in the cancer registry (Table 1). The ICD-O-3 codes were then used to further classify diagnoses and group patients into malignancy categories outlined by the BCCSS.[6] Additionally, for brain tumor patients, ICD-O-3 topography for central nervous system (CNS) locations were used to mitigate misclassification based on primary pathologic diagnosis (i.e. intracranial mixed germ cell tumors).

| Risk stratification

The cancer registry captures the first course of treatment based on chart review by a certified cancer registrar in accordance with the FORDS Manual.[35] Exposures are reported as dichotomous (Yes/No) for surgery, diagnostic biopsy, radiation, chemotherapy, hormonal therapy, immunotherapy, other, palliative, and transplant. Based on these exposures and the primary diagnosis classification, risk strata were constructed from the BCCSS system (Table 1).[6]

| Follow-up definitions

The institutional cancer registry provided the base cohort for all childhood cancer diagnoses. These registry data were merged, using MRN and a durable key unique patient identifier, with EHR data through the Duke Enterprise Data Unified Content Explorer (DEDUCE) to extract all visits in the PHO and PNO clinic encounters to identify eligible patients. To determine appropriate follow-up, all visits in the PBMT clinic and the Duke Cancer Institute were also extracted. Inadequate follow-up was defined as a survivor not being seen during the five to seven year window after the date of initial diagnosis.

| Spatial Variables

DEDUCE was also used to export the longitude and latitude coordinates of the home address, zip code, and the census block group Federal Information Processing Standards (FIPS) code for each survivor. Using ArcGIS 10.5.1 (ESRI, Redlands, CA), we calculated the Euclidean (straight line) distance from the address of each survivor to the nearest COG-affiliate site[36] in North Carolina (NC), South Carolina (SC), and Virginia (VA). Analysis was limited to survivors whose coordinates were in NC, SC, and VA. Additionally, using spatial point-in-polygon joining operations, we identified the zip code-level Rural-Urban Commuting Area (RUCA) codes and the block group-level Area-Deprivation Index (ADI) for each survivor. RUCA is a categorical classification for rural vs urban areas that takes into account population density and distance to nearest urban centers. ADI is an indexed composite of seventeen variables related to social determinants of health from the United States Census and American Community Survey that captures socioeconomic disadvantage at the census block group level.[37,38] A high ADI percentile, which represents greater disadvantage, has been shown to correlate with a number of adverse health outcomes.[39,40]

| Statistical analyses

Patients were grouped according to whether or not they were seen in a Duke Cancer Clinic five to seven years after initial date of diagnosis. Patient characteristics were compared between those that were seen in this window versus those that were not seen. Using the Cancer Registry, we utilized the last known date of contact to ensure that patients survived through the five to seven year window after their initial date of diagnosis before including them for analysis. Continuous variables are presented as medians (standard deviations), and differences were compared using the t-test. Categorical variables are presented as counts (proportions). Differences were compared using the χ^2test. For all analyses, risk strata were categorized as a three level categorical variable (low, intermediate, and high risk).

Logistic regression was used to estimate the association between follow-up and risk stratification both in bivariate analyses and after adjusting for known covariates including ALL indicator, gender, age at diagnosis, race, and indicator of local state of residence. Local state of residence was defined as residing in NC, SC or VA to minimize potential confounding associations between risk strata, distance from medical center, and follow-up care. Because our primary variable of interest consisted of three levels (i.e. risk stratification), we utilized a multiple degree of freedom lack of fit test to compare a baseline model where risk stratification was excluded and separately, a model where it was included.

Subsequent models that included risk stratification and an indicator of local state of residence were used to determine if that association varied by a broad geographic indicator. Bonferroni adjustments were made for multiple comparisons. Only patients with complete data for all covariates were included for each analysis, and effective sample sizes are included for all tables and figures. In reviewing the correlation among all predictors in our models, we found no evidence that suggested multicollinearity might be an issue. All statistical analyses were conducted R version 3.6.1.

| RESULTS

3.1 | Cancer Registry and Childhood Cancer Survivorship Cohort Construction

A total of 3,117 patients less than or equal to eighteen years of age at diagnosis were captured by the institutional cancer registry between January 1, 1994 and November 30, 2012. Of these patients, 2,206 were considered analytic cases in the cancer registry and included in our analysis. Based on ICCC-3 coding, 1312 were pediatric oncology patients and 894 were pediatric neuro-oncology patients, of whom 871 and 546 were seen in PHO or PNO clinic during their estimated on treatment window, respectively. For patients not seen in either clinic, a manual chart review of a 10% random sample of patients revealed that nearly all patients were either referred for transplant (hematopoietic stem cell or solid organ) after receipt of initial therapy at an outside institution or they were treated exclusively by another department (i.e. dermatology, gynecology-oncology, medical oncology or surgical oncology). For the PNO cohort, 274 patients died with or without documented relapse (30.6%) and ninety-seven (10.9%) had disease recurrence without documented death during the seven-year follow-up period, for a total of 275 PNO event-free survivors. For the PHO cohort, 312 patients died with or without documented relapse (23.4%) and 172 (13.1%) had disease recurrence without documented death during the seven-year follow-up period, for a total 590 PHO event-free survivors included in the analytic cohort (Figure 1).

3.2 | Disease Classifications and Risk Stratification

ICD-O-3 codes and ICCC-3 were used to further group patients to align with the BCCSS. In the final analytic survivorship cohort, there were 225 patients with leukemia, 275 with CNS tumors, and 365 patients with extracranial solid tumor or lymphoma. Based on the BCCSS risk stratification system,[6] there were 191, 496, and 158 low, intermediate, and high risk survivors, respectively.

3.4 | Follow-up Care Analysis

Of the 865 survivors included in the final analytic cohort, 282 (32.6%) were not seen in any oncology-related subspecialty clinic (PHO, PNO, PBMT, Medical Oncology) five to seven years after the initial diagnosis.

Factors associated with follow-up included risk strata (0.001), age (p=0.008), primary diagnosis (p<0.001), and race/ethnicity (p=0.010). Risk strata was the primary association of interest for this analysis. Low risk patients yielded the highest follow-up with 78% of patients, followed by high risk patients at 70%, and intermediate risk patients at 63%. To include spatial data elements, we restricted analysis to survivors within NC, SC, and VA (n=787). Similar associations with follow-up patterns were observed as well as lower ADI national percentile (p=0.011) and distance to primary treatment site (p<0.0001), though not distance to a COG-affiliated site (p=0.729) (Table 2). Manual chart review of a 10% randomized sample of patients lost to follow-up from each risk strata (n=37) revealed that most patients were instructed to follow-up; however there was no additional follow-up documented in EHR. One low risk, four intermediate risk, and no high risk patients had documentation of care transferred to another institution.

We first built a logistic model to test the unadjusted association between risk stratification and appropriate follow-up care through a likelihood ratio test (LRT). There was strong evidence (p<0.001) that there was an association between risk stratification and follow-up care (Table 3). Pairwise comparisons showed that the odds of receiving follow-up in the five to seven year window after initial diagnosis in the intermediate risk strata was half the odds of receiving follow-up in the low risk strata (OR 0.482; CI 0.3, 0.774, p<0.001). However, there is insufficient evidence to suggest the odds of follow-up is different in pairwise comparisons between high and intermediate risk and high and low risk survivors (p=0.41 and p=0.23, respectively). We then built a multiple logistic regression model adjusted for diagnosis of ALL, gender, age at diagnosis, race/ethnicity as potential confounders. This attenuated the observed association and, after controlling for potential confounding, there was insufficient evidence to suggest there is an association between risk strata and follow-up care (p=0.17) (Table 4).

To test the hypothesis that risk strata may have a different effect if the patient is closer to the primary treatment center (i.e. Duke), we created a model that included the interaction between local residence and risk strata and used a lack of fit test to look for evidence that living in NC, SC or VA modified the effect of risk strata. There was insufficient evidence (p=0.14) to suggest that "local" patients modified the risk strata effect on likelihood to follow-up in the five to seven year window. For the survivorship cohort limited to NC, SC, and VA, we then constructed a multiple logistic regression model to adjust for ALL, gender, age at diagnosis, race/ethnicity as well as ADI, distance to primary treatment center, distance to COG-affiliated site, and RUCA. This yielded similar results with attenuation of the initially observed associated between follow-up care and risk strata, thus there was insufficient evidence to suggest an association between risk strata and follow-up care in the five to seven year window after the initial diagnosis (p=0.11) (Table 4).

4 | DISCUSSION

Nearly a third of survivors were not seen in a subspecialty clinic five to seven years after their initial diagnosis and there was insufficient evidence to support an association between risk strata for serious adverse health outcomes and likelihood of follow-up. This retrospective cohort study of survivors demonstrates the feasibility and utility of cancer registry data to construct a single-institution childhood cancer survivorship cohort. Furthermore, the integration of EHR and spatial data to expand analysis to include outcomes beyond mortality and disease recurrence, such as appropriate follow-up care, holds great promise to serve as a platform to enhance clinical oncology research with real-world data. The similar proportions of survivors seen in a subspecialty clinic between five and seven years, regardless of their risk stratification, is striking. Both the high proportion of patients not seen in clinic and the implementation of a validated risk-stratification system for survivors highlight the need for improved systems to enhance retention of patients for follow-up care.

While literature to support the benefits of long-term follow-up for survivors is abundant[2,24,41] and guidelines inform clinicians on surveillance of late effect,[11] a uniform definition of "appropriate" follow-up and "lost to" follow-up remains elusive. One limitation with the measurement of the primary "lost to follow-up" outcome variable in our cohort is whether or not patients seen in clinic were truly lost to follow-up or if they had moved or transferred care to another tertiary care center. On manual review of a random sample of thirty-seven patients lost to follow-up revealed only five had documentation of transfer to another institution. A

CCSS analysis of 6,176 survivors showed approximately 40.3% self-reported survivor-focused care within the preceding two years at the baseline questionnaire, which then declined to 30.2% within at the most recent follow-up questionnaire.[42] This markedly differs from the findings in our cohort with 67% of survivors with a subspecialty clinic visit between five and seven years after the initial date of diagnosis. The difference in these results may be due to a longer length of time since initial diagnosis at the time of the baseline questionnaire, with a mean of 17.5 years from diagnosis,[43] and the exclusion of survivors <18 years in the CCSS compared to our cohort. Other single-institution[25] and regional[44]studies for predictors of follow-up care reported higher rates of follow-up for younger patients and leukemia survivors, similar to our cohort, though definitions of appropriate follow-up were not uniform.

The opportunity to implement findings from larger cohort studies, such as the risk stratification system developed by the BCCSS,[4-6,45] helps to frame the concept of "appropriate" follow-up care. The heterogeneity of survivors, based on their primary diagnosis and treatment exposures, merits careful consideration clinically on an individual level as well as a health systems level to optimize the follow-up care for this diverse population. Risk-adapted long-term follow-up care from recent European guidelines[3] and the role of onco-primary care to facilitate transition of care for low, and potentially intermediate, risk survivors to primary care providers through the use of survivorship care plans offers a strategy to improve care. Indeed, the initial association between low risk survivors and increased likelihood of follow-up was attenuated by adjustment for primary diagnosis of ALL. Prioritization of efforts to target high risk survivors to re-engage them in subspecialty survivorship care is essential.

Routine follow-up in a survivorship clinic serves as an initial step; however, refinement of therapy-related exposures and ascertainment of adherence to guideline recommendations is critical to ensure timely delivery of appropriate care. Dichotomous exposures for chemotherapy, radiation, and surgery may lead to misclassification of patients into lower risk strata, despite receipt of high doses of chemotherapy known to significantly increase risk for late effects (i.e. cumulative anthracycline exposure and risks for cardiotoxicity).[46] Over the past twenty years, significant progress in the risk prediction of late effects based on cumulative doses of specific chemotherapy agents and radiation[20-23] calls for the inclusion of more granular exposure data to further sharpen risk stratification models. Successful de-escalation of therapy in the last several decades with sustained improvement in survival, such as leukemia and lymphoma,[47] the observed reduction in the burden of late effects based on decade of treatment[48], and the emergence of novel agents with unknown late effects[7,8]necessitate an adaptable system to incorporate more than simply dichotomous treatment exposures for risk stratification.

Potential disparities in appropriate follow-up care were identified in this analysis by race/ethnicity and SES. Specifically survivors who identified as black or Hispanic, were from locations with a higher area deprivation index, were of older age or lived further away from the primary treatment center or COG-affiliated site were less likely to receive follow-up care five to seven years after the initial diagnosis. Multivariable analyses were not executed for these associations, as the primary predictor focused on risk stratification, thus interpretation of these observed associations must be made with caution. Furthermore, the complex interactions of race/ethnicity and SES in general,[49,50] and in childhood cancer survivorship research,[51] are often difficult to disentangle. The development of potential interventions to reengage at-risk patients through an integrated implementation science and community-based participatory research approach could help ensure health equity for all survivors.

Future applications for biomedical informatics tools, such as the integration of cancer registry and EHR data, include extraction of cumulative chemotherapy exposures from the EHR as well as discrete data elements to assess adherence to guideline recommendations based on specific treatment exposures. Validation of new risk stratification models based on more granular data with large cohorts, such as the CCSS, may help bolster new models of care. Single-institution cohorts, particularly with lack of sufficient follow-up time since the widespread adoption of the modern EHR, may display limited power to detect rare events, yet may be sufficiently versatile to translate research through an implementation science approach. Furthermore, collaboration with other institutions within the NCDB with uniform cancer registry data and similar EHR

platforms is feasible to maximize the impact of real-world data. The Cancer Moonshot prioritized the creation of innovative oncology data-sharing, the reduction of health disparities, and the identification of health-care delivery models to optimize care for survivors.[52] The National Cancer Institute also recently launched the Childhood Cancer Data Initiative to foster collaboration between researchers and build an infrastructure to integrate data from multiple sources.[53] This study provides a reproducible model to integrate cancer registry and EHR data to construct risk-adapted survivorship cohorts to assess follow-up care and aligns with national efforts to apply biomedical informatics methods to revolutionize clinical research and improve survivorship care.

CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest.

ACKNOWLEDGEMENTS

Funding: David Noyd is supported by NIH "Transfusion Medicine and Hematology" (5T32 HL007057-44) training grant effective 07/01/2019.

REFERENCES

1. Jemal A, Ward EM, Johnson CJ, et al. Annual Report to the Nation on the Status of Cancer, 1975-2014, Featuring Survival. *J Natl Cancer Inst.* 2017;109(9):djx030.

2. Landier W, Skinner R, Wallace WH, et al. Surveillance for Late Effects in Childhood Cancer Survivors. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2018;36(21):2216-2222.

3. Michel G, Mulder RL, van der Pal HJH, et al. Evidence-based recommendations for the organization of long-term follow-up care for childhood and adolescent cancer survivors: a report from the PanCareSurFup Guidelines Working Group. *J Cancer Surviv.* 2019.

4. Wallace WH, Blacklay A, Eiser C, et al. Developing strategies for long term follow up of survivors of childhood cancer. *BMJ (Clinical research ed).* 2001;323(7307):271-274.

5. Edgar AB, Duffin K, Borthwick S, Marciniak-Stepak P, Wallace WH. Can intensity of long-term follow-up for survivors of childhood and teenage cancer be determined by therapy-based risk stratification? *BMJ Open.* 2013;3(8).

6. Frobisher C, Glaser A, Levitt GA, et al. Risk stratification of childhood cancer survivors necessary for evidence-based clinical long-term follow-up. *Br J Cancer.* 2017;117(11):1723-1731.

7. Moslehi JJ. Cardiovascular Toxic Effects of Targeted Cancer Therapies. *New England Journal of Medicine.*2016;375(15):1457-1467.

8. Wanchoo R, Karam S, Uppal NN, et al. Adverse Renal Effects of Immune Checkpoint Inhibitors: A Narrative Review. *Am J Nephrol.*2017;45(2):160-169.

9. Byun DJ, Wolchok JD, Rosenberg LM, Girotra M. Cancer immunotherapy - immune checkpoint blockade and associated endocrinopathies. *Nat Rev Endocrinol.* 2017;13(4):195-207.

10. Chow EJ, Antal Z, Constine LS, et al. New Agents, Emerging Late Effects, and the Development of Precision Survivorship. *Journal of Clinical Oncology.* 2018;36(21):2231-2240.

11. Group CsO. *Long-Term Follow-Up Guidelines for Survivors of Childhood, Adolescent and Young Adult Cancers, Version 5.0.* Monrovia, CA2018.

12. Kremer LCM, Mulder RL, Oeffinger KC, et al. A worldwide collaboration to harmonize guidelines for the long-term follow-up of childhood and young adult cancer survivors: a report from the International Late Effects of Childhood Cancer Guideline Harmonization Group. *Pediatric blood & cancer.* 2013;60(4):543-549.

13. Armenian SH, Hudson MM, Mulder RL, et al. Recommendations for cardiomyopathy surveillance for survivors of childhood cancer: a report from the International Late Effects of Childhood Cancer Guideline Harmonization Group. *The Lancet Oncology.* 2015;16(3):e123-e136.

14. Mulder RL, Kremer LCM, Hudson MM, et al. Recommendations for breast cancer surveillance for female survivors of childhood, adolescent, and young adult cancer given chest radiation: a report from the International Late Effects of Childhood Cancer Guideline Harmonization Group. *The Lancet Oncology.* 2013;14(13):e621-e629.

15. Skinner R, Mulder RL, Kremer LC, et al. Recommendations for gonadotoxicity surveillance in male childhood, adolescent, and young adult cancer survivors: a report from the International Late Effects of Childhood Cancer Guideline Harmonization Group in collaboration with the PanCareSurFup Consortium. *The Lancet Oncology.*2017;18(2):e75-e90.

16. Dorp Wv, Mulder RL, Kremer LCM, et al. Recommendations for Premature Ovarian Insufficiency Surveillance for Female Survivors of Childhood, Adolescent, and Young Adult Cancer: A Report From the International Late Effects of Childhood Cancer Guideline Harmonization Group in Collaboration With the PanCareSurFup Consortium. *Journal of Clinical Oncology.* 2016;34(28):3440-3450.

17. Clement SC, Kremer LCM, Verburg FA, et al. Balancing the benefits and harms of thyroid cancer surveillance in survivors of Childhood, adolescent and young adult cancer: Recommendations from the international Late Effects of Childhood Cancer Guideline Harmonization Group in collaboration with the PanCareSurFup Consortium. *Cancer Treat Rev.* 2018;63:28-39.

18. Clemens E, van den Heuvel-Eibrink MM, Mulder RL, et al. Recommendations for ototoxicity surveillance for childhood, adolescent, and young adult cancer survivors: a report from the International Late Effects of Childhood Cancer Guideline Harmonization Group in collaboration with the PanCare Consortium. *The Lancet Oncology.*2019;20(1):e29-e41.

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. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.*2009;27(14):2308-2318.

20. Clark RA, Mostoufi-Moab S, Yasui Y, et al. 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). *The Lancet Oncology.* 2020;21(3):436-445.

21. Kovalchik SA, Ronckers CM, Veiga LH, et al. Absolute risk prediction of second primary thyroid cancer among 5-year survivors of childhood cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2013;31(1):119-127.

22. Chow EJ, Chen Y, Hudson MM, et al. Prediction of Ischemic Heart Disease and Stroke in Survivors of Childhood Cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2018;36(1):44-52.

23. Chow EJ, Chen Y, Kremer LC, et al. Individual prediction of heart failure among childhood cancer survivors. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2015;33(5):394-402.

24. Signorelli C, Wakefield CE, Fardell JE, et al. The impact of long-term follow-up care for childhood cancer survivors: A systematic review. *Crit Rev Oncol Hematol.* 2017;114:131-138.

25. May L, Schwartz DD, Frugé E, et al. Predictors of Suboptimal Follow-up in Pediatric Cancer Survivors. *Journal of pediatric hematology/oncology.* 2017;39(3):e143-e149.

26. Hashibe M, Kirchhoff AC, Kepka D, et al. Disparities in cancer survival and incidence by metropolitan versus rural residence in Utah.*Cancer Med.* 2018;7(4):1490-1497.

27. Singh GK, Williams SD, Siahpush M, Mulhollen A. Socioeconomic, Rural-Urban, and Racial Inequalities in US Cancer Mortality: Part I-All Cancers and Lung Cancer and Part II-Colorectal, Prostate, Breast, and Cervical Cancers. *J Cancer Epidemiol.* 2011;2011:107497-107497.

28. Weaver KE, Geiger AM, Lu L, Case LD. Rural-urban disparities in health status among US cancer survivors. *Cancer.*2013;119(5):1050-1057.

29. Zahnd WE, Davis MM, Rotter JS, et al. Rural-urban differences in financial burden among cancer survivors: an analysis of a nationally representative survey. *Supportive care in cancer : official journal of the Multinational Association of Supportive Care in Cancer.*2019;27(12):4779-4786.

30. Walling EB, Fiala M, Connolly A, Drevenak A, Gehlert S. Challenges Associated With Living Remotely From a Pediatric Cancer Center: A Qualitative Study. *J Oncol Pract.* 2019;15(3):e219-e229.

31. Fluchel MN, Kirchhoff AC, Bodson J, et al. Geography and the burden of care in pediatric cancers. *Pediatric blood & cancer.*2014;61(11):1918-1924.

32. Alvarez E, Keegan T, Johnston EE, et al. Adolescent and young adult oncology patients: Disparities in access to specialized cancer centers. *Cancer.* 2017;123(13):2516-2523.

33. Boffa DJ, Rosen JE, Mallin K, et al. Using the National Cancer Database for Outcomes Research: A Review. *JAMA Oncology.*2017;3(12):1722-1728.

34. Steliarova-Foucher E, Stiller C, Lacour B, Kaatsch P. International Classification of Childhood Cancer, third edition. *Cancer.*2005;103(7):1457-1467.

35. Cancer Co. Facility Oncology Registry Data Standards 2016. In: American College of Surgeons; 2016.

36. Group CsO. Children's Oncology Group Locations. https://www.childrensoncologygroup.org/index.php/locations/. Published 2020. Accessed August 5, 2020.

37. Kind AJH, Buckingham WR. Making Neighborhood-Disadvantage Metrics Accessible - The Neighborhood Atlas. *N Engl J Med.*2018;378(26):2456-2458.

38. Singh GK. Area deprivation and widening inequalities in US mortality, 1969-1998. *Am J Public Health.* 2003;93(7):1137-1143.

39. Lantos PM, Maradiaga-Panayotti G, Barber X, et al. Geographic and Racial Disparities in Infant Hearing Loss. *Otolaryngol Head Neck Surg.* 2018:194599818803305.

40. Durfey SNM, Kind AJH, Buckingham WR, DuGoff EH, Trivedi AN. Neighborhood disadvantage and chronic disease management. *Health Serv Res.* 2019;54 Suppl 1(Suppl 1):206-216.

41. Nekhlyudov L, Mollica MA, Jacobsen PB, Mayer DK, Shulman LN, Geiger AM. Developing a Quality of Cancer Survivorship Care Framework: Implications for Clinical Care, Research, and Policy. *JNCI: Journal of the National Cancer Institute.* 2019;111(11):1120-1130.

42. Casillas J, Oeffinger KC, Hudson MM, et al. Identifying Predictors of Longitudinal Decline in the Level of Medical Care Received by Adult Survivors of Childhood Cancer: A Report from the Childhood Cancer Survivor Study. *Health Serv Res.* 2015;50(4):1021-1042.

43. Oeffinger KC, Mertens AC, Sklar CA, et al. Chronic health conditions in adult survivors of childhood cancer. *N Engl J Med.*2006;355(15):1572-1582.

44. Welch JJG, Kenney LB, Hirway P, et al. Understanding predictors of continued long-term pediatric cancer care across the region: A report from the Consortium for New England Childhood Cancer Survivors. *Pediatr Blood Cancer.* 2017;64(10).

45. Eiser C, Absolom K, Greenfield D, et al. Follow-up after childhood cancer: evaluation of a three-level model. *Eur J Cancer.*2006;42(18):3186-3190.

46. Bates JE, Howell RM, Liu Q, et al. Therapy-Related Cardiac Risk in Childhood Cancer Survivors: An Analysis of the Childhood Cancer Survivor Study. *Journal of Clinical Oncology.* 2019;37(13):1090-1101.

47. Hudson MM, Neglia JP, Woods WG, et al. Lessons from the past: Opportunities to improve childhood cancer survivor care through outcomes investigations of historical therapeutic approaches for pediatric hematological malignancies. *Pediatric Blood & Cancer.*2012;58(3):334-343.

48. Gibson TM, Mostoufi-Moab S, Stratton KL, et al. Temporal patterns in the risk of chronic health conditions in survivors of childhood cancer diagnosed 1970-99: a report from the Childhood Cancer Survivor Study cohort. *Lancet Oncol.* 2018;19(12):1590-1601.

49. Shavers VL. Measurement of socioeconomic status in health disparities research. *J Natl Med Assoc.* 2007;99(9):1013-1023.

50. Cheng TL, Goodman E. Race, Ethnicity, and Socioeconomic Status in Research on Child Health. *Pediatrics.* 2015;135(1):e225-e237.

51. Kehm RD, Spector LG, Poynter JN, Vock DM, Altekruse SF, Osypuk TL. Does socioeconomic status account for racial and ethnic disparities in childhood cancer survival? *Cancer.* 2018;124(20):4090-4097.

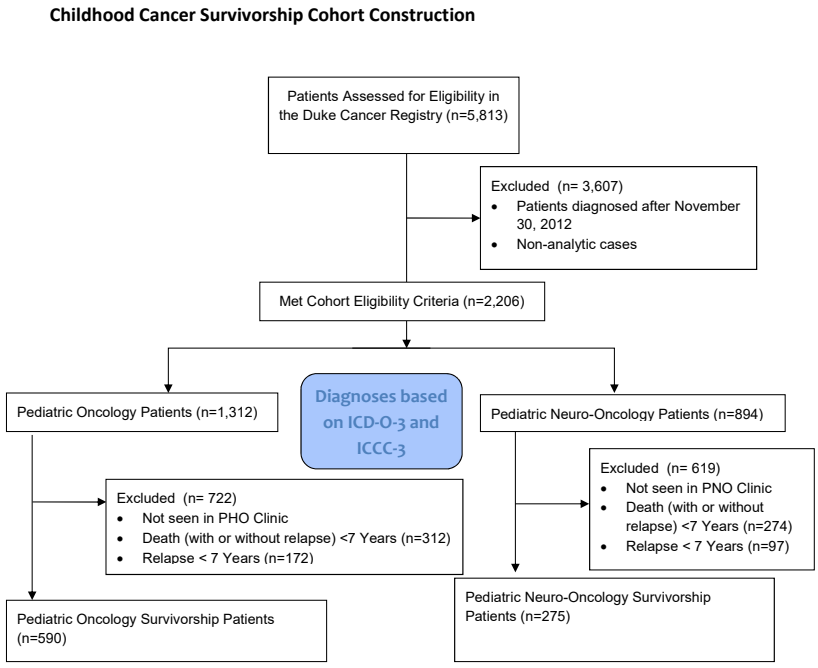
52. Jaffee EM, Dang CV, Agus DB, et al. Future cancer research priorities in the USA: a Lancet Oncology Commission. *The Lancet Oncology.* 2017;18(11):e653-e706.

53. Institute NC. Childhood Cancer Data Initiative. https://www.cancer.gov/research/areas/childhood/childhood-cancer-data-initiative. Accessed March 22, 2020.

Data available on request due to privacy/ethical restrictions: The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

LEGENDS

FIGURE 1. Construction of final analytic childhood cancer survivorship cohort based on inclusion and exclusion criteria from cancer registry data for patients diagnosed between January 1, 1994 and November 30, 2012.



Hosted file

Table 1.pdf available at https://authorea.com/users/371237/articles/489652-integration-of-cancer-registry-and-electronic-health-record-data-to-construct-a-childhood-cancer-survivorship-cohort-facilitate-risk-stratification-and-assess-appropriate-follow-up-care

Hosted file

Table 2.pdf available at https://authorea.com/users/371237/articles/489652-integration-of-cancer-registry-and-electronic-health-record-data-to-construct-a-childhood-cancer-survivorship-cohort-facilitate-risk-stratification-and-assess-appropriate-follow-up-care

Hosted file

Table 3.pdf available at https://authorea.com/users/371237/articles/489652-integration-of-cancer-registry-and-electronic-health-record-data-to-construct-a-childhood-cancer-survivorship-cohort-facilitate-risk-stratification-and-assess-appropriate-follow-up-care

Hosted file

`Table 4.pdf` available at https://authorea.com/users/371237/articles/489652-integration-of-cancer-registry-and-electronic-health-record-data-to-construct-a-childhood-cancer-survivorship-cohort-facilitate-risk-stratification-and-assess-appropriate-follow-up-care