HHS Public Access

Author manuscript

Semin Oncol. Author manuscript; available in PMC 2022 April 25.

Published in final edited form as:

Semin Oncol. 2020 February ; 47(1): 56–64. doi:10.1053/j.seminoncol.2020.02.006.

Using big data in pediatric oncology: Current applications and future directions

Ajay Major, MD, MBA[a], **Suzanne M. Cox, PhD, MPH**[b], **Samuel L. Volchenboum, MD, PhD**[c,*]

[a]Section of Hematology Oncology, University of Chicago Medicine, Chicago, IL

[b]Biological Sciences Division, University of Chicago, Chicago, IL

[c]Pediatric Hematology Oncology, University of Chicago, Chicago, IL

Abstract

Pediatric cancer is a rare disease with a low annual incidence, which presents a significant challenge in being able to collect enough data to fuel clinical discoveries. Big data registry trials hold promise to advance the study of pediatric cancers by allowing for the combination of traditional randomized controlled trials with the power of larger cohort sizes. The emergence of big data resources and data-sharing initiatives are becoming transformative for pediatric cancer diagnosis and treatment. This review discusses the uses of big data in pediatric cancer, existing pediatric cancer registry initiatives and research, the challenges in harmonizing these data to improve accessibility for study, and building pediatric data commons and other important future endeavors.

Keywords

Pediatric oncology; Pediatric cancer; Big data; Data sharing; Data science; Informatics

Introduction

Pediatric cancer is relatively rare, with an estimated 11,000 new cases diagnosed among children from birth to 14 years of age in the United States in 2019 [1]. Although the mortality rate for this age group has declined by 65 percent from 1970 to 2016, cancer remains the leading cause of death among children, with about 1,200 children expected to die in the United States in 2019 from cancer [1]. If adolescents are included, then the total number of new cancer diagnoses (from birth to 19 years of age) is more than 14,500 children and young adults per year in the United States [1, 2]. In stark contrast, there were 1.7 million new cases of adult cancer in the United States in 2016 [3]. Thus, while each

*Corresponding author. Pediatric Hematology Oncology, University of Chicago, 900 E 57th St, 5th flr, Chicago, IL 60637. slv@uchicago.edu (S.L. Volchenboum).

Conflicts of interest

Suzanne M. Cox: **Consulting or Advisory Role:** Litmus Health. Samuel L. Volchenboum: **Stock and Other Ownership Interests:** Litmus Health. **Consulting or Advisory Role:** Accordant Health Services, a CVS Caremark Company.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi: 10.1053/j.seminoncol.2020.02.006.

diagnosis and death of a child or adolescent from cancer represents an unfathomable tragedy for a family, pediatric cancer remains a rare disease with a low annual incidence. Even the most common types of pediatric cancer, including leukemias, central nervous system tumors, and lymphomas, present a significant challenge in being able to collect enough data to fuel clinical discoveries [1].

There are obvious benefits to improved evidence-based guidelines for the management of pediatric cancers, as childhood cancers are successfully treated in approximately 80% of cases in high-income countries [4]. This reduction in mortality can be attributed to advances in diagnosis, risk stratification, better supportive care, and treatment. Given the paucity of childhood cancer cases and the challenge in amassing sufficient numbers of patients for meaningful study, big data registry trials hold promise to advance the study of pediatric cancers by allowing for the combination of traditional randomized controlled trials (RCTs) with the power of larger cohort sizes. The emergence of big data resources and data-sharing initiatives are becoming transformative for pediatric cancer diagnosis and treatment. This review discusses the uses of big data in pediatric cancer, existing pediatric cancer registry initiatives and research, the challenges in harmonizing these data to improve accessibility for study, and building pediatric data commons and other important future endeavors.

Big data in pediatric cancer

Big data has been defined as information assets “characterized by such a high volume, velocity and variety to require specific technology and analytical methods for its transformation into value” [5, 6]. In oncology, big data has many possible components including biological, clinical, and administrative data about cancer patients in both structured and unstructured formats. These data can be mined to answer questions about genomics, outcomes, and therapeutic efficacy [7].

Big data have been extensively utilized in adult oncology research due to the relatively large number of patients and the availability of large connected patient registries [8]. Yet, the use of big data in pediatric cancer is not as well-developed. The National Cancer Institute (NCI) has spearheaded efforts to promote collaborative use of big data across oncology subspecialties, including the creation of the Childhood Cancer Data Initiative, a $50 million initiative with an “aggressive focus” on data sharing that was announced in 2019 [9]. This initiative is exciting for big data advocates, as coordinated data-sharing initiatives can enable the pooling of patient-level data for children across the country, and in some cases across the globe, to accelerate the pace of and possibility for discovery. While the creation of large, collaborative data commons in pediatric cancer is in the early stages, there already exist many smaller patient registries that are disease-, society- or geography-specific, which can lay the foundation for robust data sharing and discovery [10].

There are a number of existing data-sharing initiatives that include pediatric cancer data, sponsored by various sub-discipline interest groups, national health agencies, and academic consortia. Key US-based data sources and initiatives are summarized in Table 1. Many of these were detailed in an earlier publication, and therefore the focus here is on updates and new initiatives [10].

The NCI has combined a number of data resources into a **Cancer Research Data Commons (CRDC) Ecosystem** [11]. The goal of the ecosystem is to create data commons nodes and sources to co-locate data storage and computing infrastructure with services, tools, and applications [12]. Within the ecosystem are several data sources, including the Therapeutically Applicable Research To Generate Effective Treatments (TARGET), the Proteomics Data Commons (PDC), the Imaging Data Commons, the Genomic Data Commons (GDC), the Integrated Canine Data Commons, and the Human Tumor Atlas Network, among others that are emerging. Currently, the GDC is the only data source in the CRDC that includes pediatric cancer data.

The TARGET consortium is a collaboration of investigators comprised mainly of members of the Children's Oncology Group (COG), a clinical trials group devoted exclusively to childhood and adolescent cancer research [13]. TARGET researchers collaborate with the COG to access clinical expertise and biospecimens across the network, with the goal of producing genomics data that will facilitate molecular discoveries and aid translation of those findings into effective therapies. TARGET includes data from children with acute lymphoblastic leukemia (ALL), acute myeloid leukemia, kidney tumors, neuroblastoma, and osteosarcoma [14, 15].

The GDC centralizes, standardizes, and makes accessible data from large-scale NCI programs such as TARGET and its adult equivalent, The Cancer Genome Atlas [16, 17]. Within the GDC, genomics data and associated clinical data can be stored and analyzed, allowing researchers to compare findings across studies. One of the key GDC initiatives has been to harmonize the NCI's cancer genomics data, including both processing the genomic data with uniform pipelines as well as developing a data model with uniform terms and definitions for biospecimens and clinical data.

The Surveillance, Epidemiology, and End Results (SEER) Program is a registry sponsored by the National Cancer Institute that collects cancer incidence and survival data from 19 U.S. geographic areas and represents approximately 34% of the United States population [18]. SEER includes data on all ages and has been used extensively for childhood cancer research.

The **Children's Oncology Group (COG)** is the largest international pediatric oncology research consortium and executes both late-phase and early-phase clinical trials through an extensive clinical trial network in conjunction with the NCI. The COG network is made up of over 150 academic medical centers in the US, Canada, and Australia. Most children in the US are treated on or according to a COG trial. COG has maintained several large patient registries, including the Childhood Cancer Research Network and Project: EveryChild, which represent greater than 90% of pediatric oncology cases under age 15 in the United States [19]. In addition, the COG and the NCI established a prospective precision medicine trial in 2017, entitled Molecular Analysis for Therapy Choice (Pediatric MATCH), with the goal of building a genomic registry to study early-phase therapies in pediatric solid tumors [20].

The **Pediatric Cancer Data Commons** (PCDC) at the University of Chicago works with stakeholders to create data commons for pediatric cancer data [21]. Through an iterative consensus-building process, the PCDC team creates and ballots consensus data dictionaries for pediatric cancer. So far, the PCDC has created stable versions of international data dictionaries for neuroblastoma, rhabdomyosarcoma, germ cell tumors, and acute myelogenous leukemia. The PCDC team negotiates data sharing and use agreements with the data owners, permitting the harmonization and movement of data into the commons for subsequent sharing and analysis. As of January 2020, the PCDC contains data on over 25,000 children and is growing rapidly. By using a common COG identifier assigned to every child, data in the PCDC can be linked in real time to other sources, including genomic data in TARGET and tissue availability in the Nationwide Children's Biopathology Center. An expansion of the PCDC is underway, as new diseases are being added rapidly, including acute lymphocytic leukemia, acute myelogenous leukemia, Hodgkin lymphoma, osteosarcoma, Ewing sarcoma, and germ cell tumors. In addition to facilitating the harmonization of data from completed trials, the consensus data dictionaries will be used to power the next generation of clinical trials. The PCDC team works closely with the National Cancer Institute to update the NCI thesaurus [22] and caDSR [23] repositories, allowing clinical trialists to select balloted elements for future data collection forms. The PCDC has been built to be part of the Cancer Research Data Commons Ecosystem. By leveraging the Gen3 technical infrastructure, developed for the Genomic Data Commons, the PCDC will more easily integrate and interoperate with the other data commons nodes in the CRDC.

The **Treehouse Childhood Cancer Initiative** is a registry of gene expression data from over 11,000 pediatric tumor samples that is freely available to researchers for comparative genomic analysis and the development of novel targeted therapies [24]. Nine hospital and consortia partners have shared gene expression data on tumors along with patient-privacy protected clinical data, including age, gender, and disease type. These data are available for public download.

St. Jude Children's Hospital has developed a large cloud-based pediatric cancer registry called SJCARES that was specifically designed for low- and middle-income countries to enable international collaborative research [25]. St. Jude also sponsors a large source of international genomic data called the PeCan Data Portal, which provides "interactive visualizations of pediatric cancer mutations across various projects at St. Jude Children's Research Hospital and its collaborators" [26]. PeCan currently has data from 4,877 patients with 23 different diagnoses and over 88,000 mutations [26]. Most recently, St. Jude Cloud was launched, which houses high-throughput genomics data from patients at St. Jude in a searchable interface with advanced analytic tools [27, 28].

The **Gabriella Miller Kids First Pediatric Research Program** is a cloud-based pediatric genomics registry that aims to examine the genetic causes of childhood cancer and structural birth defects and to advance personalized medicine for the detection, therapy, and the management of these diseases in children [29]. The Kids First Data Resource is distinguished from other data commons by its goal to understand the genetic causes and links between childhood cancer and structural birth defects.

The **National Program of Cancer Registries**, which is sponsored by the Centers for Disease Control and Prevention (CDC), supports state-based cancer registries in the United States and represents approximately 97% of the United States population [30]. In 2014, the NPCR established a Pediatric and Young Adult Early Case Capture program to register pediatric cancer cases within 30 days of diagnosis [31].

Pediatric cancer registry-based research

To determine how childhood cancer registries are currently being used in pediatric oncology research, a literature review was performed using the PubMed database and "pediatric cancer" and "registry" search terms between years 2018–2020 [32]. The search returned 751 articles; the authors reviewed all articles and determined 214 articles to be relevant to the present review. Studies were included if they presented research that utilized a multi-site registry for the study of pediatric or adolescent and young adult (AYA) patients with malignant tumors. Studies were excluded if they only presented research on benign tumors or if they only used single-institution data sources/registries. Each study was coded by its dominant research domain, which were selected *a priori*. The overall results in terms of the number of publications in each domain are summarized in Figure 1. The full list of publications reviewed are available in the Appendix.

This review strategy is limited in several ways including: very recent publication time frame, limited to a single publication database, and inclusion only of research and registries that have resulted in a published peer-reviewed paper. It is important to note that other important registries exist and have had an impact on the study of pediatric cancer. Rather this literature review serves as a summary of recent activity with the aim of showing trends in publication and potential gaps in the field.

Descriptive epidemiologic research

The largest domain consisted of epidemiological research that used registry data retrospectively to describe pediatric cancer in terms of risk stratification, prognosis, health outcomes, and health services research. The SEER database was prominently used in many studies, including by research groups not based in the United States, to study a wide range of outcomes. For example, using SEER, Deng et al found that several prognostic factors, including age, use of radiotherapy, and gross total resection were associated with improved survival in pediatric pineoblastomas [33].

Many of the registries performing epidemiologic research were country-based, such as the Finnish Cancer Registry, Canadian Cancer Registry, French National Registry of Childhood Solid Tumors, and the Italian Neuroblastoma Registry. There were also several multinational consortium registries represented, including the International Society of Paediatric Oncology (SIOP), Cooperative Weichteilsarkom Studiengruppe, and European Rhabdoid Registry, as well as the large bone marrow transplant research groups CIBMTR, EBMT, ABMTRR, JSHCT, and the Asia-Pacific Blood and Marrow Transplantation Group which represents 18 countries. The EUROCOURSE group, a network of regional or national cancer registries in 12 countries in Southern and Eastern Europe, produced several studies on Wilms tumor

and neuroblastoma to demonstrate the power of pooling individual patient data from a large region, as well as to demonstrate inferior outcomes in the region [34].

The ability to pool data into a large data commons is a key feature of many of these consortium-sponsored studies, and would also benefit smaller registries. For example, the Cancer Registry of New Caledonia, a French territory in the Pacific Ocean, described 162 total cases of pediatric cancer between 1994 and 2012 [35]. Pooling these cases into a larger multinational registry would help with comparison of New Caledonia to other nations, as well as enable the registry to participate in larger pediatric cancer trials.

One prospective observational registry, the Thai Acute Leukemia Working Group, part of the Thai Society of Hematology, was identified. This registry assessed survival in AYA *v*adult patients with acute lymphocytic leukemia (ALL), and found improved survival in patients who underwent stem cell transplantation [36]. The ability to conduct prospective epidemiologic research is a highlight of registries and data commons, although the majority of research continues to be retrospective with the exception of large treatment protocol-generating groups such as SIOP and COG.

Several studies used large registries for the study of rare pediatric cancers, such as solid pseudopapillary pancreatic tumors with the Italian Pediatric Rare Tumor Registry [37] or Merkel cell carcinoma with SEER [38]. For rare tumors, it is clear that large datasets can improve research quality; a single-center registry in Kerala, India, of pediatric brain tumors was unable to achieve statistical significance due to small sample size, and specifically called for the establishment of a brain tumor registry for the Indian population [39].

Survivorship research

The second most-represented domain was survivorship research, which used registry data to study late effects, secondary malignancies, and toxicities in survivors of childhood cancer. There were several large registries that produced the majority of this research, including the Swiss Childhood Cancer Registry, the Nordic Adult Life after Childhood Cancer in Scandinavia group, and PanCareSurFup. The PanCareSurFup registry includes an impressive cohort of over 83,0 0 0 5-year survivors of childhood cancer diagnosed from 1940 to 2011, pooled from registries in 12 European countries [40]. Based on this cohort, PanCareSurFup has released comprehensive evidence-based guidelines for childhood cancer survivors, including surveillance for breast cancer, cardiovascular disease, infertility, and thyroid cancer [40].

Two prospective registry trials of childhood cancer survivors included the Utah Cancer Registry and the Swedish Childhood Cancer Registry. Examples of studies include focus group research to assess follow-up preferences in survivorship [41] and an internet-based cognitive behavioral therapy intervention for psychological distress [42].The ability to both retrospectively and prospectively recruit patients from these registries for survivorship studies is a highlight of data commons.

The limits of the literature review are evident in the failure to identify publications from key registries of survivor data. For example, the St. Jude LIFE study of survivors of pediatric

cancer has relevant studies such as a description of health outcomes, quality of life, and social attainment among adult survivors of neuroblastoma [43]. Also from St. Jude, the Childhood Cancer Survivor Study (CCSS) has resulted in a number of studies on survivors of childhood cancer, such as research on the risk of thyroid cancer, factors associated with physical activity, and neurocognitive concerns [44–46].

Genomics research

Several studies used pediatric cancer registries to perform genomic research including pharmacogenomics, germline predisposition, and genetic risk scoring. The PanCareLIFE registry, which represents childhood cancer survivors across Europe treated with cisplatin, carboplatin, or cranial radiotherapy, performed genotyping to assess for gene variants associated with ototoxicity [47] and infertility [48]. The TARGET database was also used in one study by Kim et al in conjunction with the International Pleuropulmonary Blastoma/DICER1 Registry to study for the presence of germline DICER1 mutations in various types of pediatric cancer [49].

New registry description

The literature review also identified several newly-established pediatric cancer registries, including a new SIOP registry for Wilms tumor [50] and the Pediatric Proton Consortium Registry [51]. Several new registries focused specifically on the needs of low- and middle-income countries, including a Caribbean cancer registry hub described by Spence et al [52] and a real-time pediatric cancer case reporting system entitled VIGICANCER in Columbia described by Ramirez et al [53].

Data harmonization

Several studies specifically presented data harmonization initiatives in pediatric cancer registries, including a study by the European Society for Blood and Marrow Transplantation (EBMT), which developed a universal system for benchmarking of outcomes in stem cell transplantation reporting among EBMT members [54]. There were also several studies that specifically addressed disparities for smaller registries, including Mazzucco et al, which compared the Palermo Province Cancer Registry with larger Italian and Southern European cancer registries. The authors found comparable cancer incidence and survival in pediatric populations [55], arguing that smaller registries can play a vital role in pediatric cancer surveillance of local communities in conjunction with large population-based registries. Thus, the need for data harmonization to allow integration and comparison across data sources is critical.

Prospective registry-based therapeutic trials

Registry-based randomized clinical trials (RRCT) are an emerging type of study design in oncology, used more commonly in cardiology, which enable rapid patient recruitment and randomization through a previously-established registry to study interventions in real-world populations [56]. Although the literature review did not identify any RRCTs for pediatric cancer patients, there were several prospective registry-based trials studying specific therapeutic interventions. Yamasaki et al performed a prospective multicenter

registry study with 30 institutions of pediatric patients with medulloblastoma, stratifying patients to one of several treatment protocols based on their risk profiles and assessing outcomes [57]. There were also several prospective, non-randomized single-center registry trials of radiotherapy for rhabdomyosarcoma [58, 59] and craniopharyngioma [60], although single-center registries were not included in the literature review.

Other research types

The literature review also found smaller numbers of studies on pediatric registries used for radiology research, supportive and palliative care, hereditary pediatric cancer syndromes, EHR integration, and consensus guidelines (see Appendix). Notably, there was a prospective registry established by the National Cancer Institute entitled Adolescent and Young Adult Health Outcomes and Patient Experiences which recruited AYA patients to a large multicenter registry study of care quality and psychosocial outcomes throughout treatment and survivorship [61].

Cross-disciplinary big data collaborations

Several instances of collaborative registry development between oncology groups and other professions or disciplines were found. For instance, at least two studies devoted to pediatric oncology nutrition were derived from data in the Swiss Childhood Cancer Registry [62–64]. A recent review of big data for pediatric oncology nutrition highlights some of the main sources of data for research (claims data, EHR data) [65]. A group of pediatric oncology radiologists formed a photon/proton consortium with the aim of collecting and standardizing data [66]. Additionally, there were a significant number of studies describing long-term toxicities in childhood cancer survivors, which would be relevant to primary care, general practitioners, and cardiologists, but there is little registry-based data on how to maximize follow-up and long-term surveillance of toxicities. The survivor registries at St. Jude may provide a model for other data commons efforts [67, 68].

Technical challenges and barriers to data sharing

Despite the wide variety of disparate registries identified in the literature review, there are clear efforts by large consortia to implement intra-registry data harmonization. For example, one of the studies utilized Delphi consensus methodology to harmonize the numerical units of busulfan reporting in stem cell transplant registries [69]. However, there were no reports of studies that intended to harmonize data between registries. Rather, the review identified multiple registries that already exist within larger geographic or disease-specific international registries. While the case has been made that regional registries are redundant and national registries are more cost-effective [70], improved harmonization and creation of data commons would allow these regional registries, such as the Palermo Province Cancer Registry identified in our literature review, to participate in clinical trials on the international stage [55].

The goal should always be to collect standardized data from children with cancer and then connect sources of disparate data. Where standardized data are not available, they may be harmonized to a common standard after collection, as is now common practice. An example

is the collection of clinical, biospecimen, and genomic data for a child with cancer. The clinical data from COG can be harmonized to a common standard and made available in the Pediatric Cancer Data Commons, the standardized genomic data available in the Gabriella Miller Kids First Data Resource Center or the Genomic Data Commons, and the information about biospecimen availability in the Biopathology Center (BPC) at Nationwide Children's Hospital. Connecting these three sources of data are the Universal Specimen Identifier assigned by the BPC. The user can then query for a specific cohort of patients and immediately see for which patients genomic data and biospecimens are available. Without rigorous data standardization and harmonization processes, this would not be possible.

There are many challenges in collecting, aggregating, and using big data for pediatric oncology research, many similar across all areas of medicine. For instance, rapid reporting for children's cancer remains difficult and error-prone, hindering fast dissemination of results [71]. Using genomic data remains difficult, as highlighted by the Treehouse Project group, and includes issues such as data location, characterization, quality assessment, use approval, and compliance [72]. Leveraging data from electronic health records (EHR) is especially challenging owing to customized implementations and a general lack of adherence to data standards [73]. Aggregating clinical data remain extremely challenging, largely due to lack of standardization and especially the rampant use of free-text notes. Identifying cohorts of children for oncology studies requires custom development of criteria [74]. Variations in EHR implementations and rampant customization of the installations continue to plague efficient sharing of data for research [75]. The MIRACUM consortium in Germany has started to tackle the issues of data quality and suggests a framework for a unified, standardized, and harmonized EHR for clinical research. Similarly, the PEDSnet national network strives to be a platform for discovery for pediatric research [76], but it remains to be seen whether or not standardization of data components will exceed a small handful of clinical features.

Data sharing agreements remain a significant barrier to effective and timely research. A recent survey of academic researchers revealed that rather than the legal team being a bottleneck for data use agreement execution, the problems mostly stemmed from procedural inefficiencies, incomplete information, a lack of incentives and familiarity with academic practices, and unresponsiveness of faculty [77].

Patient consent issues remain a major factor in the use of data for research. Most consents for pediatric studies are paper-based and stored as scanned files or as printed and bound pages. Often, documentation of consent for specific studies or tests will require manual abstraction from the paper consent into an electronic data capture system. For maximal use of data, patients should be consented for broad use of their data, and the consent should be captured as granular, computable data. Informed consent for children is also complicated by the need to be re-consented when patients reach the age of majority.

Future directions

The literature review identified many types of pediatric cancer registries, from small single-institutions to large multinational consortia. The wide diversity of registry studies published

in the past 2 years indicate a growing interest in the field of pediatric oncology to answer both clinical and basic science questions with large datasets. However, it is clear from the literature review that the landscape of big data in pediatric oncology is heavily siloed, with small and large registries collecting disparate types of data with poor inter-registry harmonization and lack of data standardization across the field. Further, the literature review demonstrates disparities in pediatric cancer registries for developing nations and resource-poor economies, with some nascent efforts to better integrate these patients into clinical trials. Many opportunities exist to help solve these problems and improve care for children with cancer.

Prospective clinical trials and developing novel therapies

There is a considerable lag time in approval of oncologic therapies for children, estimated at a 6.5 year median interval between first-in-human and first-in-child drug approvals by the FDA [78]. Off-label prescribing in children approaches 90% [79]. As such, there is growing pressure to increase the throughput of clinical trials in pediatric oncology, including hastening real-time registration of pediatric cancer patients in registries.

However, the literature review demonstrated that the vast majority of pediatric cancer registry studies over the past several years have been retrospective, with very few prospective studies and even fewer prospective registry trials to study novel therapies or treatment protocols. Although there have been previous attempts to use observational data to replicate RCTs in adults, there are very few of these studies in pediatrics or pediatric oncology, and their methodological accuracy has not been verified [80]. As such, prospective protocol trials and RCTs remain the standard in studying new therapies; however, only the largest international pediatric registries have historically generated this kind of research, including COG (North American), SIOP, and Cooperative Weichteilsarkom Studiengruppe (European) consortia.

The implementation of RRCTsmay enable a larger group of registries, including small or geographically-isolated registries, to participate in clinical trials [56]. Further, with appropriate integration of real-time registration systems, these registries may also enable more rapid clinical trial enrollment to expedite the study of novel therapies.

Although clinical trial enrollment of pediatric patients has been increasing over time [81], there are barriers among parents [82] and investigators [83] towards clinical trial enrollment. For parents, important factors influencing participation in a clinical trial include trust, appropriate timing, a transparent discussion of the risks and benefits, and a motivation for children to participate [82]. For providers, barriers include a lack of awareness of available trials, risk of studies, distance to a site that could provide the trial, and the time required to discuss the trial with parents [82, 83].

The EU Paediatric Regulation [84], the United States' Pediatric Research Equity Act (PREA), the Best Pharmaceuticals for Children Act [85], and the recent Paediatric Investigation Plan [86] all serve to provide incentives for the development of drugs with pediatric indications. But successful clinical trials in children require the collection of

robust, high-quality research data. Standardization of data collection across studies and disease types will massively enable pediatric cancer research.

Pediatric cancer in developing nations

An encouraging finding was several publications from smaller registries in developing nations, with several groups specifically calling for the creation of larger registries to promote pediatric cancer research. There are clearly disparities in collection and use of big data for childhood cancer in low- and middle-income nations.

These articles highlight the burden of pediatric cancer in lower- and middle-income countries where incidence rates are higher than in higher-income countries (384,000 *v*45,000 new cases per year), and 5-year survival rates are much lower (30% *v*80%) [4, 87]. Additionally, efforts are underway to attempt to predict the global burden and amount of undiagnosed cases [88]. The pediatric cases of cancer in low- and middle-income countries are not included in cancer registries [89]. There are several high-impact opportunities in developing countries to improve diagnosis and treatment of children with cancer by leveraging data. For example, the International Network for Cancer Treatment and Research cancer registry program is coordinating an African Cancer Registry Network [90]. To make real progress in big data-driven research, data collection and ‘an accurate appraisal of the global burden of childhood cancer is a necessary first step’ [4]. Big data-based registry protocols are already being used in developing nations, for example a 2001 SIOP Wilms tumor protocol being used in southern India [91]. The logistics of enrolling patients to a registry may present a significant barrier in some developing nations. The literature review found an example of how this barrier to entry can be lowered, in a study by Shen et al in which the Pediatric Oncology Network Database, a web-based patient registration system designed by St. Jude Children’s, was used to enroll pediatric patients with ALL onto a clinical trial of a treatment protocol [92]. Overall, low- and middle-income countries represent an enormous potential for improvement in pediatric cancer diagnosis and survival. Harnessing big data techniques to aid in the collection, storage, and study of data from these countries will enable data-driven methods for risk stratification and better treatment outcomes.

The data commons

Data commons are essential to the study of pediatric cancer big data. A data commons is a repository of harmonized data, available in a cloud-based infrastructure, with accompanying tools to support analysis [93]. Transparent governance policies should help define how researchers can interact with the data. As outlined previously by our group, the steps necessary to build a cancer data commons include gaining support for developing the platform, creating and balloting an international data standard, and harmonizing retrospective data for inclusion in the commons [10]. Further, the stable, balloted data dictionary will be used to update the NCI’s thesaurus as well as the caDSR, thus allowing the standardized terms to be incorporated into data collection forms for subsequent studies.

The Pediatric Cancer Data Commons remains the only multi-disease international clinical trials data commons for pediatric cancer. The PCDC team works with clinical disease

leaders to develop an international consortium and then simultaneously execute data sharing agreements and iteratively developing a balloted, consensus data dictionary. Data are then transformed prior to incorporation into the commons, and any new terms are harmonized with the NCI's thesaurus and caDSR. A standardized process is defined to allow researchers to apply for line-level data access to facilitate subsequent analyses. Outcome data is periodically updated, requiring reports and updates from the consortia data managers. To truly be characterized as a data commons, the platform must link the clinical data to other data sources through a common identifier, the COG USI in the case of the PCDC. Finally, analytic tools should be available to the user, allowing cloud-based studies and obviating the need for data download.

The PCDC is one of a growing number of data commons that can be part of the NCI's Cancer Research Data Commons ecosystem, a cloud-based infrastructure with interactive portals that give users access to data and facilitate in-depth data analyses [94]. Currently, the main nodes in the CRDC include the Genomic Data Commons [95], the Proteomic Data Commons [96], the Integrated Canine Data Commons, the Imaging Data Commons, and the Human Tissue Atlas Network data coordinating center. Other data sources include the Gabriella Miller Kids First Data Resource Center [97] and the St. Jude Genomics Cloud [98]. New initiatives like the recently-launched Center for Cancer Data Harmonization [98, 99] and the upcoming Cancer Data Aggregator will serve as platforms/resources to help the data nodes connect and harmonize data, while assisting end users in data preparation and submission.

Conclusions

The volume and variety of data available for pediatric cancer research is increasing quickly. The amount of genomic data generated annually is projected to exceed 2 exabytes, perhaps up to 40 exabytes [100]. To effectively harness this torrent of data for research requires integration of multiple kinds of data, including clinical information. Currently, most data collection is siloed and performed with little regard for data standardization, leaving valuable troves of information isolated and often useless for subsequent integration and analysis. But the landscape is changing for the better. The use of common data standards when developing data collection tools is increasing, and there is a growing appreciation among researchers and clinicians that data collected as discrete, standardized elements will be much more easily harmonized with other sources of data and thus be much more useful for research.

Outlined here is the landscape of big data for pediatric cancer research. The growing number of interconnected repositories for these data represent an enormous opportunity for clinicians and researchers. Data standardization will play a key role in ensuring interoperability of data. The maturing NCI Cancer Research Data Ecosystem - especially the Cancer Research Data Commons - provides an environment for harmonized data collection, remote storage, aggregation of disparate data, and cloud-based tools for analysis. There remains much work to be done to enable the tools and services that will guarantee that the diverse kinds of data will be usable for research, including broad consents, more easily implemented data sharing and use agreements, and cooperation among data stakeholders to

agree on standardized means of data capture. This is an incredibly exciting time for pediatric oncology research, and the availability of high-quality data will be transformative for the field, and hopefully lead to new ways to diagnose and improve the treatment of children with cancer.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

[1]. Childhood Cancers. National Cancer Institute 2015. https://www.cancer.gov/types/childhood-cancers (Accessed 31 October 2019).

[2]. Pediatric and Young Adult Early Case Capture | CDC 2019. https://www.cdc.gov/cancer/npcr/early-case-capture.htm (Accessed 31 October 2019).

[3]. Surveillance, epidemiology, and end results program n.d. https://seer.cancer.gov/ (Accessed 4 March 2017).

[4]. Bhakta N, Force LM, Allemani C, et al. Childhood cancer burden: a review of global estimates. Lancet Oncol 2019;20:e42–53 10.1016/s1470-2045(18)30761-7. [PubMed: 30614477]

[5]. De Mauro A, Greco M, Grimaldi M. A formal definition of Big Data based on its essential features. &ctx_ver=Z 2016;39:88.

[6]. ACCP Journals. American College of Clinical Pharmacology n.d. https://accp1.onlinelibrary.wiley.com/doi/full/10.1002/jcph.1141 (Accessed 20 January 2020).

[7]. Chambers DA, Amir E, Saleh RR, et al. The impact of big data research on practice, policy, and cancer care. Am Soc Clin Oncol Educ Book 2019;39:e167–75. [PubMed: 31099675]

[8]. Andrade J, Cox SM, Volchenboum SL. Large-scale data sharing initiatives in genomic oncology. Adv Mol Pathol 2018;1:135–48 https://doi.org/. doi: 10.1016/j.yamp.2018.06.009.

[9]. The childhood cancer data initiative: sharing for progress. National Cancer Institute 2019. https://www.cancer.gov/news-events/cancer-currents-blog/2019/lowy-childhood-cancer-data-initiative (Accessed 20 January 2020).

[10]. Volchenboum SL, Cox SM, Heath A, Resnick A, Cohn SL, Grossman R. Data commons to support pediatric cancer research. Am Soc Clin Oncol Educ Book 2017;37:746–52. [PubMed: 28561664]

[11]. NCI Cancer Research Data Commons | CBIIT n.d. https://datascience.cancer.gov/data-commons (Accessed 20 January 2020).

[12]. Grossman RL. Progress toward cancer data ecosystems. Cancer J 2018;24:126–30. [PubMed: 29794537]

[13]. Therapeutically Applicable Research to Generate Effective Treatments. NCI Office of Cancer Genomics. 2013. https://ocg.cancer.gov/programs/target (Accessed 20 January 2020).

[14]. Ma X, Liu Y, Liu Y, et al. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. Nature 2018;555:371–6. [PubMed: 29489755]

[15]. TARGET Data Matrix. Office of Cancer Genomics n.d. https://ocg.cancer.gov/programs/target/data-matrix (Accessed 21 January 2020).

[16]. Newly launched Genomic Data Commons to facilitate data and clinical information sharing. National Institutes of Health (NIH); 2016. https://www.nih.gov/news-events/news-releases/newly-launched-genomic-data-commons-facilitate-data-clinical-information-sharing. (Accessed January 21, 2020).

[17]. Home | NCI Genomic Data Commons n.d. https://gdc.cancer.gov/ (Accessed 21 January 2020).

[18]. About the SEER Program - SEER. SEER n.d. https://seer.cancer.gov/about/overview.html (Accessed 20 January 2020).

[19]. O'Leary M, Krailo M, Anderson JR, Reaman GH, Group Children's Oncology. Progress in childhood cancer: 50 years of research collaboration, a report from the Children's Oncology Group. Semin Oncol 2008;35:484–93. [PubMed: 18929147]

[20]. NCI-COG Pediatric MATCH. National Cancer Institute 2015. https://www.cancer.gov/about-cancer/treatment/clinical-trials/nci-supported/pediatric-match (Accessed 20 January 2020).

[21]. Pediatric Cancer Data Commons – Connect. Share. Cure n.d. http://commons.cri.uchicago.edu (Accessed 21 January 2020).

[22]. NCI Thesaurus n.d. https://ncithesaurus.nci.nih.gov/ncitbrowser/ (Accessed 21 January 2020).

[23]. CDE Browser 5.3.5 n.d. https://cdebrowser.nci.nih.gov/cdebrowserClient/deBrowser.html (Accessed 21 January 2020).

[24]. Bjork I, Peralez J, Haussler D, Spunt SL, Vaske OM. Data sharing for clinical utility. Cold Spring Harb Mol Case Stud 2019:5 10.1101/mcs.a004689.

[25]. SJCARES Registry n.d. https://www.stjude.org/global/sjcares/registry.html (Accessed 20 January 2020).

[26]. St. Jude PeCan Data Portal n.d. https://pecan.stjude.org (Accessed 27 March 2017).

[27]. Home. St Jude Cloud n.d. https://www.stjude.cloud/ (Accessed 21 January 2020).

[28]. Home. St Jude Cloud n.d. https://www.stjude.cloud/ (Accessed 21 January 2020).

[29]. Seven Bridges announces the launch of largest pediatric data resource as a member of the Gabriella Miller Kids First Data Resource Center. Seven Bridges 2018. https://www.sevenbridges.com/press/releases/gabriella-miller-kids-first-data-resource-center/ (Accessed 4 November 2019).

[30]. About NPCR | Cancer | CDC 2019. https://www.cdc.gov/cancer/npcr/about.htm (Accessed 20 January 2020).

[31]. Pediatric and Young Adult Early Case Capture | CDC 2019. https://www.cdc.gov/cancer/npcr/early-case-capture.htm (Accessed 20 January 2020).

[32]. pubmeddev. Home - PubMed - NCBI n.d. www.ncbi.nlm.nih.gov/pubmed (Accessed 21 January 2020).

[33]. Deng X, Yang Z, Zhang X, et al. Prognosis of Pediatric Patients with Pineoblastoma: A SEER Analysis 1990–2013. World Neurosurg 2018;118:e871–9. [PubMed: 30031180]

[34]. Doganis D, Panagopoulou P, Tragiannidis A, et al. Survival and mortality rates of Wilms tumour in Southern and Eastern European countries: Socioeconomic differentials compared with the United States of America. Eur J Cancer 2018;101:38–46. [PubMed: 30014973]

[35]. Hartmann E, Missotte I, Dalla-Pozza L. Cancer Incidence Among Children in New Caledonia, 1994 to 2012. J Pediatr Hematol Oncol 2018;40:515–21. [PubMed: 30044351]

[36]. Limvorapitak W, Owattanapanich W, Utchariyaprasit E, et al. Better survivals in adolescent and Young adults, compared to adults with acute lymphoblastic leukemia - A multicenter prospective registry in Thai population. Leuk Res 2019;87:106235. [PubMed: 31675661]

[37]. Crocoli A, Grimaldi C, Virgone C, et al. Outcome after surgery for solid pseudopapillary pancreatic tumors in children: Report from the TREP project-Italian Rare Tumors Study Group. Pediatr Blood Cancer 2019;66: e27519. [PubMed: 30362240]

[38]. Paulson KG, Nghiem P. One in a hundred million: Merkel cell carcinoma in pediatric and young adult patients is rare but more likely to present at advanced stages based on US registry data. J Am Acad Dermatol 2019;80:1758–60. [PubMed: 30165170]

[39]. Govindan A, Parambil RM, Alapatt JP. Pediatric Intracranial Tumors over a 5-Year Period in a Tertiary Care Center of North Kerala, India: A Retrospective Analysis. Asian J Neurosurg 2018;13:1112–17.

[40]. Byrne J, Alessi D, Allodji RS, et al. The PanCareSurFup consortium: research and guidelines to improve lives for survivors of childhood cancer. Eur J Cancer 2018;103:238–48. [PubMed: 30286417]

[41]. Ramsay JM, Mann K, Kaul S, Zamora ER, Smits-Seemann RR, Kirchhoff AC. Follow-up care provider preferences of adolescent and young adult cancer survivors. J Adolesc Young Adult Oncol 2018;7:204–9. [PubMed: 29346008]

[42]. Woodford J, Wikman A, Cernvall M, et al. Study protocol for a feasibility study of an internet-administered, guided, CBT-based, self-help intervention (ENGAGE) for parents of children previously treated for cancer. BMJ Open 2018;8:e023708.

[43]. Wilson CL, Brinkman TM, Cook C, et al. Clinically ascertained health outcomes, quality of life, and social attainment among adult survivors of neuroblastoma: A report from the St. Jude Lifetime Cohort. Cancer 2020 10.1002/cncr.32678.

[44]. Gramatges MM, Morton LM, Yasui Y, et al. Telomere Length-Associated Genetic Variants and the Risk of Thyroid Cancer in Survivors of Childhood Cancer: A Report from the Childhood Cancer Survivor Study (CCSS). Cancer Epidemiol Biomarkers Prev 2019;28:417–19. [PubMed: 30377209]

[45]. Devine KA, Mertens AC, Whitton JA, et al. Factors associated with physical activity among adolescent and young adult survivors of early childhood cancer: A report from the childhood cancer survivor study (CCSS). Psycho-Oncology 2018;27:613–19 10.1002/pon.4528. [PubMed: 28805953]

[46]. Kenzik KM, Huang I-C, Brinkman TM, et al. The Childhood Cancer Survivor Study-Neurocognitive Questionnaire (CCSS-NCQ) revised: item response analysis and concurrent validity. Neuropsychology 2015;29:31–44. [PubMed: 24933482]

[47]. Clemens E, Broer L, Langer T, et al. Genetic variation of cisplatin-induced ototoxicity in non-cranial-irradiated pediatric patients using a candidate gene approach: The International PanCareLIFE Study. Pharmacogenomics J 2019 10.1038/s41397-019-0113-1.

[48]. van der Kooi A-LLF, Clemens E, Broer L, et al. Genetic variation in gonadal impairment in female survivors of childhood cancer: a PanCareLIFE study protocol. BMC Cancer 2018;18:930. [PubMed: 30257669]

[49]. Kim J, Schultz KAP, Hill DA, Stewart DR. The prevalence of germline DICER1 pathogenic variation in cancer populations. Mol Genet Genomic Med 2019;7:e555. [PubMed: 30672147]

[50]. Vujanić GM, Gessler M, Ooms AHAG, et al. The UMBRELLA SIOP-RTSG 2016 Wilms tumour pathology and molecular biology protocol. Nat Rev Urol 2018;15:693–701. [PubMed: 30310143]

[51]. Hess CB, Indelicato DJ, Paulino AC, et al. An update from the pediatric proton consortium registry. Front Oncol 2018;8:165. [PubMed: 29881715]

[52]. Spence D, Argentieri MA, Andall-Brereton G, Anderson BO, Duggan C, Bodkyn C, et al. Advancing cancer care and prevention in the Caribbean: a survey of strategies for the region. Lancet Oncol 2019;20:e522–34. [PubMed: 31395471]

[53]. Ramirez O, Aristizabal P, Zaidi A, Ribeiro RC, Bravo LEVIGICANCER Working Group. Implementing a Childhood Cancer Outcomes Surveillance System Within a Population-Based Cancer Registry. J Glob Oncol 2018;4:1–11.

[54]. Snowden JA, Saccardi R, Orchard K, et al. Benchmarking of survival outcomes following haematopoietic stem cell transplantation: A review of existing processes and the introduction of an international system from the European Society for Blood and Marrow Transplantation (EBMT) and the Joint Accreditation Committee of ISCT and EBMT (JACIE). Bone Marrow Transplant 2019 10.1038/s41409-019-0718-7.

[55]. Mazzucco W, Cusimano R, Mazzola S, et al. Childhood and Adolescence Cancers in the Palermo Province (Southern Italy): Ten Years (2003–2012) of Epidemiological Surveillance. Int J Environ Res Public Health 2018:15 10.3390/ijerph15071344.

[56]. Foroughi S, Wong H-L, Gately L, et al. Re-inventing the randomized controlled trial in medical oncology: The registry-based trial. Asia Pac J Clin Oncol 2018;14:365–73. [PubMed: 29947051]

[57]. Yamasaki K, Okada K, Soejima T, Sakamoto H, Hara J. Strategy to minimize radiation burden in infants and high-risk medulloblastoma using intrathecal methotrexate and high-dose chemotherapy: a prospective registry study in Japan. Pediatr Blood Cancer 2020;67:e28012. [PubMed: 31544362]

[58]. Ludmir EB, Grosshans DR, McAleer MF, et al. Patterns of failure following proton beam therapy for head and neck rhabdomyosarcoma. Radiother Oncol 2019;134:143–50. [PubMed: 31005208]

[59]. Ludmir EB, Paulino AC, Grosshans DR, et al. Regional Nodal Control for Head and Neck Alveolar Rhabdomyosarcoma. Int J Radiat Oncol Biol Phys 2018;101:169–76. [PubMed: 29477293]

[60]. Ajithkumar T, Mazhari A-L, Stickan-Verfürth M, et al. Proton Therapy for Craniopharyngioma - An Early Report from a Single European Centre. Clin Oncol 2018;30:307–16.

[61]. Smith AW, Keegan T, Hamilton A, et al. Understanding care and outcomes in adolescents and young adult with Cancer: A review of the AYA HOPE study. Pediatr Blood Cancer 2019;66:e27486. [PubMed: 30294882]

[62]. Belle FN, Beck Popovic M, Ansari M, Otth M, Kuehni CE, Bochud M. Nutritional Assessment of Childhood Cancer Survivors (the Swiss Childhood Cancer Survivor Study-Nutrition): Protocol for a Multicenter Observational Study. JMIR Res Protoc 2019;8:e14427. [PubMed: 31738177]

[63]. Belle FN, Wenke-Zobler J, Cignacco E, et al. Overweight in childhood cancer patients at diagnosis and throughout therapy: a multicentre cohort study. Clin Nutr 2019;38:835–41. [PubMed: 29544999]

[64]. Childhood Cancer Registry (ChCR). Childhood Cancer Registry (ChCR) n.d. https://www.childhoodcancerregistry.ch/ (Accessed 20 January 2020).

[65]. Phillips CA, Pollock BH. Big data for nutrition research in pediatric oncology: current state and framework for advancement. J Natl Cancer Inst Monogr 2019;2019:127–31. [PubMed: 31532530]

[66]. Lawell MP, Indelicato DJ, Paulino AC, et al. An open invitation to join the Pediatric Proton/Photon Consortium Registry to standardize data collection in pediatric radiation oncology. Br J Radiol 2019:20190673. [PubMed: 31600082]

[67]. St. Jude LIFE Study n.d. https://www.stjude.org/research/initiatives/cancer-survivorship-research/st-jude-life-study.html (Accessed 21 January 2020).

[68]. Childhood Cancer Survivor Study n.d. https://ccss.stjude.org/ (Accessed 21 January 2020).

[69]. McCune JS, Quinones CM, Ritchie J, et al. Harmonization of Busulfan Plasma Exposure Unit (BPEU): A Community-Initiated Consensus Statement. Biol Blood Marrow Transplant 2019;25:1890–7. [PubMed: 31136799]

[70]. Registration of childhood cancer: moving towards pan-European coverage? Eur J Cancer 2015;51:1064–79. [PubMed: 25899984]

[71]. Puckett M, Neri A, Rohan E, et al. Evaluating early case capture of pediatric cancers in seven central cancer registries in the United States, 2013. Public Health Rep 2016;131:126–36. [PubMed: 26843678]

[72]. Learned K, Durbin A, Currie R, et al. Barriers to accessing public cancer genomic data. Scientific Data 2019;6:1–7. [PubMed: 30647409]

[73]. Meehan RA, Mon DT, Kelly KM, R, et al. Increasing EHR system usability through standards: conformance criteria in the HL7 EHR-system functional model. J Biomed Inform 2016;63:169–73. [PubMed: 27523469]

[74]. Phillips CA, Razzaghi H, Aglio T, et al. Development and evaluation of a computable phenotype to identify pediatric patients with leukemia and lymphoma treated with chemotherapy using electronic health record data. Pediatr Blood Cancer 2019;66:e27876. [PubMed: 31207054]

[75]. Bowles KH, Potashnik S, Ratcliffe SJ, et al. Conducting research using the electronic health record across multi-hospital systems: semantic harmonization implications for administrators. J Nurs Adm 2013;43:355–60. [PubMed: 23708504]

[76]. [No title] n.d. https://www.healthaffairs.org/doi/full/10.1377/hlthaff.2014.0127 (Accessed 20 January 2020).

[77]. Mello MM, Triantis G, Stanton R, Blumenkranz E, Studdert DM. Waiting for data: Barriers to executing data use agreements. Science 2020;367:150–2. [PubMed: 31919212]

[78]. Neel DV, Shulman DS, DuBois SG. Timing of first-in-child trials of FDA-approved oncology drugs. Eur J Cancer 2019;112:49–56. [PubMed: 30928805]

[79]. Hwang TJ, Tomasi PA, Bourgeois FT. Delays in completion and results reporting of clinical trials under the Paediatric Regulation in the European Union: A cohort study. PLoS Med 2018;15:e1002520. [PubMed: 29494592]

[80]. Christensen ML, Davis RL. Identifying the “Blip on the Radar Screen”: Leveraging Big Data in Defining Drug Safety and Efficacy in Pediatric Practice. J Clin Pharmacol 2018;58(Suppl 10):S86–93. [PubMed: 30248191]

[81]. Parsons HM, Penn DC, Li Q, et al. Increased clinical trial enrollment among adolescent and young adult cancer patients between 2006 and 2012–2013 in the United States. Pediatr Blood Cancer 2019;66:e27426. [PubMed: 30256525]

[82]. Greenberg RG, Gamel B, Bloom D, et al. Parents' perceived obstacles to pediatric clinical trial participation: findings from the clinical trials transformation initiative. Contemp Clin Trials Commun 2018;9:33–9. [PubMed: 29696222]

[83]. Smith B, Benjamin D, Bradley J, et al. Investigator barriers to pediatric clinical trial enrollment: Findings and recommendations from the Clinical Trials Transformation Initiative. Pediatrics 2018;142 796–796.

[84]. Global pediatric drug development. Curr Ther Res Clin Exp 2019;90:135–42. [PubMed: 31388369]

[85]. Ward RM, Kauffman R. Future of pediatric therapeutics: reauthorization of BPCA and PREA. Clin Pharmacol Ther 2007;81:477–9. [PubMed: 17375103]

[86]. Paediatric medicine: Paediatric Investigation Plan - EUPATI. EUPATI 2016. https://www.eupati.eu/clinical-development-and-trials/paediatric-medicine-paediatric-investigation-plan/.

[87]. Lam CG, Howard SC, Bouffet E, Pritchard-Jones K. Science and health for all children with cancer. Science 2019;363:1182–6. [PubMed: 30872518]

[88]. Ward ZJ, Yeh JM, Bhakta N, Frazier AL, Atun R. Estimating the total incidence of global childhood cancer: a simulation-based analysis. Lancet Oncol 2019;20:483–93. [PubMed: 30824204]

[89]. Rodriguez-Galindo C, Friedrich P, Alcasabas P, et al. Toward the Cure of All Children With Cancer Through Collaborative Efforts : Pediatric Oncology As a Global Challenge. J Clin Oncol 2015;33:3065–73. [PubMed: 26304881]

[90]. Pediatric Oncology - INCTR – International Network for Cancer Treatment and Research n.d. http://www.inctr.org/programs/pediatric-oncology/index.html (Accessed 21 January 2020).

[91]. John R, Kurian JJ, Sen S, et al. Clinical outcomes of children with Wilms tumor treated on a SIOP WT 2001 protocol in a tertiary care hospital in south India. J Pediatr Urol 2018;14 547.e1–547.e7. [PubMed: 30017606]

[92]. Shen S, Cai J, Chen J, et al. Long-term results of the risk-stratified treatment of childhood acute lymphoblastic leukemia in China. Hematol Oncol 2018;36:679–88. [PubMed: 30133806]

[93]. Grossman RL, Heath A, Murphy M, Patterson M, Wells W. A Case for Data Commons: toward data science as a service. Comput Sci Eng 2016;18:10–20. [PubMed: 29033693]

[94]. Build a National Cancer Data Ecosystem - Cancer Moonshot Recommendation. National Cancer Institute; 2018 https://www.cancer.gov/research/key-initiatives/moonshot-cancer-initiative/implementation/data-ecosystem.

[95]. Home | NCI Genomic Data Commons n.d. https://gdc.cancer.gov/ (Accessed 21 January 2020).

[96]. Proteomic Data Commons n.d. https://pdc.esacinc.com/pdc/ (Accessed 21 January 2020).

[97]. Working Together to Put Kids First n.d. https://kidsfirstdrc.org/ (Accessed 21 January 2020).

[98]. Home. St Jude Cloud n.d. https://www.stjude.cloud/ (Accessed 21 January 2020).

[99]. The Center for Cancer Data Harmonization Kicks Off Efforts | CBIIT n.d. https://datascience.cancer.gov/news-events/news/center-cancer-data-harmonization-kicks-efforts (Accessed 21 January 2020).

[100]. Stephens ZD, Lee SY, Faghri F, et al. Big data: astronomical or genomical? PLoS Biol 2015;13:e1002195. [PubMed: 26151137]

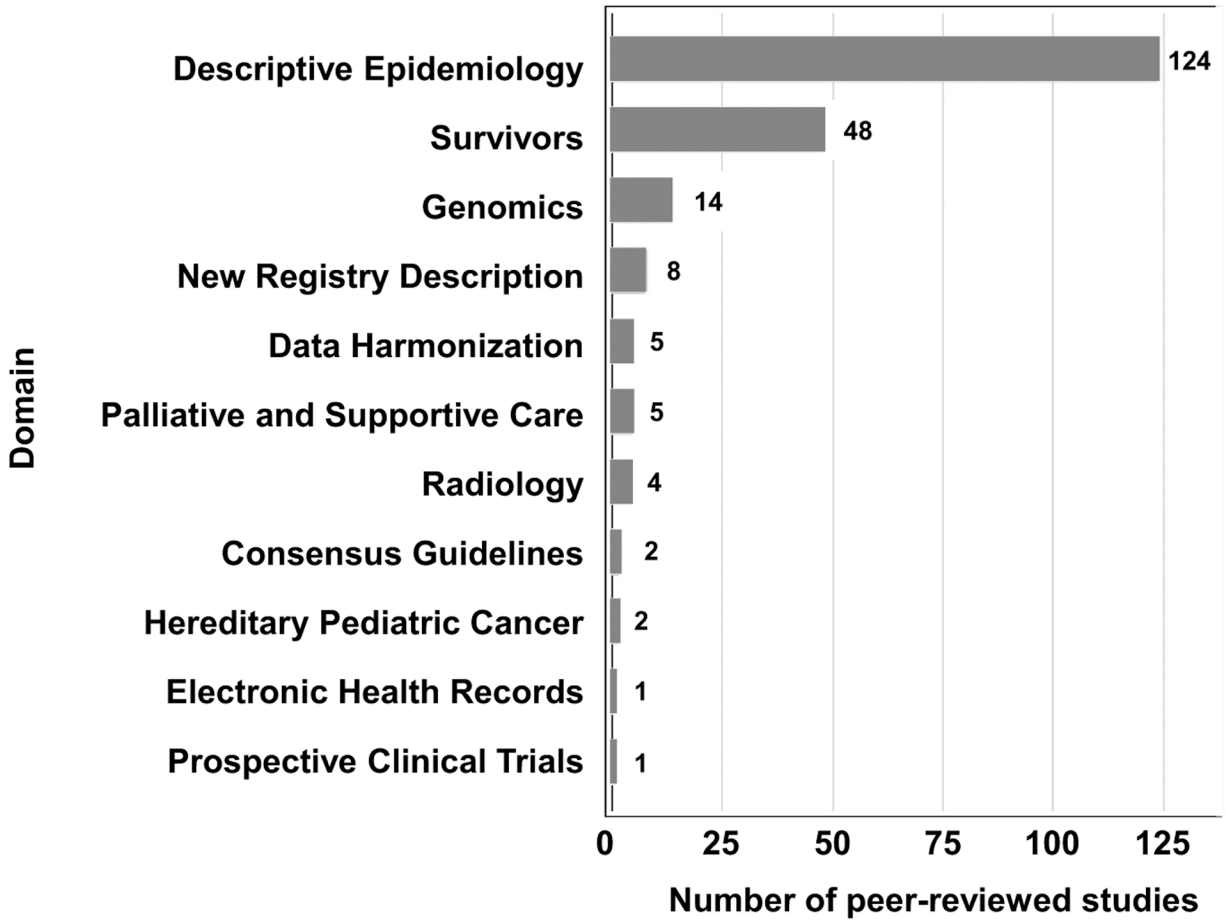

Fig. 1.
Literature review results of current (2018–2020) pediatric registry research articles by domain (N = 214).

Table 1

United States data sources and sharing initiatives that include pediatric cancer data.

Data sharing initiative	Organization	Included data source(s)	Website
Cancer Research Data Commons	National Cancer Institute	• NCI Genomic Data Commons including TARGET • Cancer Genomics Cloud Pilots • Surveillance, Epidemiology, End Results (SEER) • Proteomics, imaging, preclinical data	https://www.cancer.gov/research/nci-role/bioinformatics/cancer-research-data-ecosystem-infographic
Childhood Cancer Research Network	Children's Oncology Group	• Demographic data from children with cancer registered with COG	https://childrensoncologygroup.org/index.php/59-research/childhood-cancer-research-network-
Project: EveryChild	Children's Oncology Group	• COG registry data for children with cancer	http://projecteverychild.org/
NCI-COG Pediatric MATCH (Molecular Analysis for Therapy Choice)	Children's Oncology Group and National Cancer Institute	• Data from the NCI- and COG-sponsored precision medicine program for children with solid tumors	https://www.cancer.gov/about-cancer/treatment/clinical-trials/nci-supported/pediatric-match
Pediatric Cancer Data Commons	University of Chicago and partners	• International data from consortia representing children with neuroblastoma (International Neuroblastoma Risk Group - INRG), rhabdomyosarcoma (International Soft-Tissue Sarcoma Consortium - INSTRuCT), germ cell tumors (Malignant Germ Cell International Consortium - MaGIC), and acute myelogenous leukemia • TARGET • Nationwide Children's Biopathology Center	http://commons.cri.uchicago.edu
Treehouse Childhood Cancer Initiative	University of California, Santa Cruz	• Nine hospitals or consortia	https://treehousegenomics.soe.ucsc.edu/explore-our-data/
St. Jude Cloud	St. Jude's Hospital	• Genomic data on children diagnosed and treated at St. Jude	https://www.stjude.cloud/
PeCan Data Portal	St. Jude's Hospital	• Genomic and clinical data from patients at St. Jude and collaborators	https://pecan.stjude.cloud/
St. Jude CARES	St Jude's Hospital	• Registry data from low- and middle-income countries	https://www.stjude.org/global/sjcares/registry.html
Kids First Data Resource Center (DRC)	Gabriella Miller Kids First Pediatric Research Program	• Six partner studies, institutions, and consortia	https://kidsfirstdrc.org
National Program of Cancer Registries	Centers for Disease Control	• State-based cancer registries	https://www.cdc.gov/cancer/npcr/index.htm