HHS Public Access

Author manuscript

Paediatr Perinat Epidemiol. Author manuscript; available in PMC 2024 June 06.

Published in final edited form as:

Paediatr Perinat Epidemiol. 2018 November ; 32(6): 568–583. doi:10.1111/ppe.12519.

The International Childhood Cancer Cohort Consortium (I4C): A research platform of prospective cohorts for studying the etiology of childhood cancers

Gabriella Tikellis[a], **Terence Dwyer**[a,b], **Ora Paltiel**[c], **Gary S. Phillips**[d], **Stanley Lemeshow**[e], **Jean Golding**[f], **Kate Northstone**[f], **Andy Boyd**[f], **Sjurdur Olsen**[g], **Akram Ghantous**[h], **Zdenko Herceg**[h], **Mary H. Ward**[i], **Siri E. Håberg**[j], **Per Magnus**[j], **Jørn Olsen**[k], **Marin Ström**[g], **Somdat Mahabir**[l], **Rena R. Jones**[i], **Anne-Louise Ponsonby**[a], **Jacqueline Clavel**[m], **Marie Aline Charles**[m], **Edwin Trevathan**[n], **Zhengmin (Min) Qian**[o], **Milena M. Maule**[p], **Xiu Qiu**[q], **Yun-Chul Hong**[r], **Silvia Brandelise**[s], **Eve Roman**[t], **Melissa Wake**[a], **Jian-Rong He**[q,u], **Martha S. Linet**[v]
International Childhood Cancer Cohort Consortium

[a.]Population Epidemiology, Murdoch Children's Research Institute, Royal Children's Hospital, University of Melbourne, Melbourne, Australia

[b.]The George Institute for Global Health, University of Oxford, UK

[c.]Braun School of Public Health, Hadassah-Hebrew University Medical Center, Jerusalem, Israel

[d.]Center for Biostatistics, Department of Biomedical Informatics, Ohio State University, Columbus, Ohio, USA

[e.]Division of Biostatistics, College of Public Health, Ohio State University, Columbus, Ohio, USA

[f.]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

[g.]Centre for Fetal Programming, Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark

[h.]Epigenetics Group, International Agency for Research on Cancer, Lyon, France

[i.]Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, USA

[j.]Centre for Fertility and Health, Norwegian Institute of Public Health, Norwat

[k.]Department of Clinical Epidemiology, Aarhus University, Aarhus, Denmark

[l.]Division of Cancer Control and Population Sciences. National Cancer Institute, National Institutes of Health (NIH), Bethesda, Maryland, USA

[m.]Institut National de la Santé et de la Recherche Médicale, Centre for Research in Epidemiology and Statistics Sorbonne Paris Cité, Villejuif, France

[n.]Vanderbilt Institute for Global Health, Vanderbilt University Medical Center, Nashville, USA

[o.]College for Public Health and Social Justice, Saint Louis University, Missouri, USA

Corresponding author Gabriella Tikellis, Murdoch Children's Research Institute, The Royal Children's Hospital, Victoria, Australia, gabriella.tikellis@mcri.edu.au.

[p.]Cancer Epidemiology Unit, Department of Medical Sciences, University of Torino, Torino, Italy

[q.]Department of Woman and Child Health Care, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou, China

[r.]Institute of Environmental Medicine, College of Medicine, Seoul National University, South Korea

[s.]Boldrini Children's Center, Campinas, Brazil

[t.]Epidemiology and Cancer Statistics Group, Health Sciences, York University, UK

[u.]Nuffield Department of Women's and Reproductive Health, University of Oxford, Oxford, UK

[v.]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, USA

Abstract

Background: Childhood cancer is a rare but leading cause of morbidity and mortality. Established risk factors, accounting for less than 10% of incidence, have been identified primarily from case-control studies. However, recall, selection, and other potential biases impact interpretations particularly, for modest associations. A consortium of pregnancy and birth cohorts (I4C) was established to utilize prospective, pre-diagnostic exposure assessments and biological samples.

Methods: Eligibility criteria, follow-up methods and identification of pediatric cancer cases, are described for cohorts currently participating or planning future participation. Also described are exposure assessments, harmonization methods, biological samples potentially available for I4C research, the role of the I4C data and biospecimen coordinating centers, and statistical approaches used in the pooled analyses.

Results: Currently, six cohorts recruited over six decades (1950s – 2000s), contribute data on 388,120 mother-child pairs. Nine new cohorts from seven countries are anticipated to contribute data on 627,500 additional projected mother-child pairs within five years. Harmonized data currently includes 20+ 'core' variables, with notable variability in mother/child characteristics within and across cohorts, reflecting, in part, secular changes in pregnancy and birth characteristics over the decades.

Conclusions: The I4C is the first cohort consortium to have published findings on pediatric cancer using harmonized variables across six pregnancy/birth cohorts. Projected increases in sample size, expanding sources of exposure data (e.g., linkages to environmental and administrative databases), incorporation of biological measures to clarify exposures and underlying molecular mechanisms, and forthcoming joint efforts to complement case-control studies offer the potential for breakthroughs in pediatric cancer etiologic research.

Keywords

International Childhood Cancer Cohort Consortium (I4C); birth cohort; leukemia; childhood cancer; lifestyle factors; environmental exposures; recall bias; selection bias

INTRODUCTION

While cancer in children and adolescents is rare worldwide, it remains a leading cause of morbidity and mortality despite notable improvements in survival.[1] Established risk factors include prenatal exposure to diagnostic x-rays[2], genetic syndromes[3], and high birthweight[4] that combined, account for less than 10% of childhood cancer (CC) incidence.[5] More recently, pooled case-control studies of childhood leukemia (CL) suggest modestly increased risks associated with residential painting and pesticide use and pre-labor caesarean delivery [6,7,8] and slightly decreased risks from day care attendance, extended breastfeeding, and maternal vitamin and folic acid supplement use.[9,10] Known and suspected risk factors for CC [2] are briefly summarized in Appendix A.

Timing of exposure appears to be associated with variable CC risks, with prenatal and early postnatal periods being particularly vulnerable windows.[2,11] Increasing recognition of etiologic differences by subtype[2] underscores the need for case-control studies evaluating large numbers of distinct CC entities. While well-designed case-control studies can yield valid estimates, inherent limitations such as recall bias (differential recall of past exposures by case versus control mothers), selection bias (differential participation according to characteristics such as educational level or exposure status of cases compared with controls) and reverse causality may affect risk estimates and interpretation.

To complement and address methodologic limitations of case-control studies, pooling of multiple pregnancy/birth cohorts such as those involved in the International Childhood Cancer Consortium (I4C), could verify case-control study findings, identifying new risk factors and identify mechanisms of carcinogenesis.[12,13] Biospecimens collected prospectively are an advantage of prospective pregnancy/birth cohort studies for exploring CC etiology, although a few case-control studies have accessed archived pre-diagnostic newborn blood spots [14] or cord blood.[15]

Our objective is to report on the progress made by the I4C, furthering the description of Brown et al [16], in developing a platform through a collaborative network that provides access to repeated exposure ‘measurement’ data and biospecimens. We also describe challenges and future directions including collaborations with a consortium of case-control studies.

METHODS

Overview, structure and operations

The overarching goal of the I4C is to understand the etiology and mechanistic underpinnings of CC by exploiting prospectively collected exposure and biomarker data. The I4C *Steering Committee* includes lead investigators from cohorts, clinicians, pediatric cancer epidemiologists, molecular epidemiologists, exposure assessment experts, and funder (https://www.mcri.edu.au/research/projects/international-childhood-cancer-cohort-consortium-i4c/i4c-consortium). An international data coordinating center (IDCC) at the Murdoch Children’s Research Institute (MCRI) in Melbourne Australia houses the cohort data, manages data transfers, harmonizes variables, develops pooled

datasets and provides scientific input, and ensures the confidentiality, privacy, and security of the data. Additionally, the International Biospecimen Coordinating Center (IBCC) at the International Agency for Research on Cancer (IARC) in Lyon, France, facilitates the pooling of biological samples. The I4C projects are conducted through a*nnual open scientific meetings* and working groups attended by investigators from participating and additional emerging cohorts and other experts.

Study populations

Eligibility criteria.—Cohorts eligible for inclusion in the I4C need to recruit mothers during pregnancy or around delivery. Eligible cohorts must systematically ascertain cases of CC in the offspring and should include questionnaire and/or other exposure data that address key CC etiology-related hypotheses. The specific goals and original outcomes of the individual cohorts (e.g., pregnancy complications and /or serious chronic childhood conditions may vary, but critical data item include parental and offspring demographic, lifestyle, medical, reproductive, environmental factors, and parental occupational information. Specific responsibilities of newly joining or participating I4C cohorts include data sharing (and biospecimens- if available) for current and future proposals.

Currently contributing cohorts.—Six cohorts currently contribute data on cancer cases, exposure data and biospecimens (if available) as described in Table 1a; more details are available in the published cohort descriptions.

Data sharing

Data sharing and material transfer agreements for the I4C were developed and approved by MCRI Ethics Committee and sent to cohort investigators for approval by their Ethics Committees. Only anonymized data were requested (see Appendix B).

Follow-up methods

Strategies and time points for follow-up varied (Table 1a). Follow-up methods included postal mailings of self-administered questionnaires (ALSPAC, DNBC, MoBa), phone-administered questionnaires (DNBC, TIHS), letters to primary care physicians requesting medical records (CPP), field staff visits to extract medical record data (CPP, ALSPAC, JPP, TIHS), home visits (TIHS) and/or linkages with hospital and other national registry data (ALSPAC, DNBC, MoBa, JPS, TIHS). Follow-up response rates for the six participating cohorts were around 60-70% for most cohorts ≥7 years post-natal.

CC case ascertainment and classification

Ascertainment.—For participating cohorts, identification of CC cases has been reliant on linkage to national (ALSPAC, DNBC, MoBa and JPS) or state (TIHS) cancer registries except for CPP. The latter relied on medical records[17] and indirect methods.[18] Each potential cancer diagnosis in the CPP was reviewed by two board-certified pediatricians.

Classification.—To date, age at diagnosis for CC has been < 15 years, but going forward, will extend to < 20 years. Tumors were classified into six major groups based on the

International Classification of Diseases for Oncology (ICD-0) Third Edition.[19] For cohorts with IRB approval to access more detailed information, the following was provided: -gender, date of birth, date of diagnosis, ICD-10 code, 3-digit ICD-0-3 topographic code and 4-digit ICD-0-3 morphology code. ICD-O-3 morphology codes for leukemia included 9800-9948, gliomas 9380-9480 and lymphomas 9590-9729. From this information, the IDCC used the following six groupings: any cancer, any leukemia, acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML), any lymphomas, any central nervous system (CNS)/brain tumor, or other cancers. Due to small numbers and confidentiality issues, ALSPAC provided only any cancer, any leukemia and acute lymphoblastic leukemia. For DNBC, MoBa and JPS, mandatory reporting of cancer cases to the respective registries has been in place since the 1940s to 1960s with completeness of coverage being ≥96%.[20] In the UK, 2001 reports showed 94% of cancers ascertained during 1971–89.[21] Since CPP cases were identified through indirect methods, some cancer cases may have been missed.

Exposure data

Identification of data domains and specific variables associated with CC.—

Thirty exposure domains were established for key exposures (e.g. birthweight, folic acid supplements, and others; see Tables 3a-c). The IDCC will submit requests to obtain additional data if needed for future proposals (See Appendix C for details of process). While the main domains center around the mother and child (see Tables 3a and 3c), some information on fathers is also available (see Table 3b).

Harmonization of exposure data.—Our approach is similar other consortia.[22,23,24] Challenges include combining data from different racial and ethnic groups, collected over different time intervals or using heterogeneous data collection tools, and some variables so disparate that harmonization was not possible. The individual cohorts collected data in a standardized, structured approach from self-reported, telephone interview, or in-person administered questionnaires. Each cohort provided anonymized, individual-level. Data harmonization was carried out centrally by the IDCC project director (GT) with assistance from senior epidemiologists (TD, ALP). Each exposure variable was harmonized individually and the data evaluated for consistency within and across variables (see Appendix D)

Biological samples

Four of the participating cohorts (ALSPAC, DNBC, MoBa, TIHS) have biological specimens collected from mothers and/or offspring at various time points prior to the development of any cancer. Types of samples include: whole blood, serum, urine, and placentas from mothers; cord blood, blood (neonatal blood spots), hair, nails, and teeth from the offspring (Appendix E). All *additional emerging* cohorts are collecting a variety of biological samples.

Identification of additional emerging cohorts

Two groups of emerging cohorts are currently involved in I4C activities but not as yet contributing cancer cases, exposures or biospecimens to the pool. These are detailed in Table 1b. Group A includes five cohorts well established in recruitment and follow-up,

collecting relevant data/ biospecimens, able to ascertain CC cases, and positioned to begin contributing data to the I4C pool within the next few years: the Born in Guangzhou Cohort Study (BGCS-China), the Etude Longitudinale Française depuis l'enfance (ELFE- France), the Nascita ed Infanzia: gli Effetti dell'Ambiente (NINFEA-Italy), the Japan Environment and Children's Study (JECS-Japan) and the Korean Children's Environmental Health Study (Ko-CHENS-China). Group B consists of four cohorts in various stages of development or early recruitment and follow-up from Australia, Brazil, China and Taiwan.

Housing of Data at the IDCC: Platform, Confidentiality, Privacy and Security Measures

The data transferred to the IDCC is securely housed on a web-based application located on the MCRI's secure e-Research portal (see Appendix F). Access is restricted to authorized personnel following approval by the I4C Steering Committee and a representative from each study contributing to the pooled dataset.

For added security, data files are encrypted before being sent to the IDCC. Most studies have excluded unique personal identifiers (e.g., name, residential address) and some have excluded month and day of birth. Individuals are identified by a study-specific identification number, and additional security is provided by assigning a unique I4C identification number used as the primary identifying key. The electronic data stored at the IDCC on a secure, password protected server. The network server, web server and SQL server undergo nightly incremental backups plus a monthly full backup to tape for off-site storage. All users of the data must comply with the data sharing agreements.

Statistical consultation and support on study designs, data harmonization, and analyses

The I4C statistical team includes two senior biostatisticians (SL, GP) who provide input and advice on research proposals and undertake statistical analyses using the pooled dataset.

While complete harmonization of all questionnaire data is not feasible given cohort differences, decisions on pooling are based on the specific research question and what could be pooled with minimal compromise to the original recorded data.

Statistical methods and models used in I4C analysis

Time to event analyses use Cox proportional hazard regression models. Calculation of person-years of follow-up are based on the start time defined as the birth date (the date is set to zero years); the end time for those with cancer defined as the date of cancer diagnosis; the end time for those without cancer defined as the date the child is no longer under observation.

Statistical issues considered include: (1) accounting for different cohorts; (2) handling missing data for risk factors using multiple chained imputation techniques; (3) dealing with different lengths of follow-up of the contributing cohorts; (4) examining confounding and effect-modification of postulated risk factors; (5) finding the correct scale for continuous covariates and (6) testing the proportional hazard assumption for Cox regression models. Further details and strategies are in Appendix G.

RESULTS

Cohorts currently contributing data.

Six cohorts (Table 1a) currently contribute data on 388,120 mother-child pairs as well as less extensive paternal data for certain domains (Tables 3a-c). Recruitment periods span over six decades from the late 1950s (CPP), mid 1960s-mid70s (JPS), late 1980s (TIHS), early 1990s (ALSPAC), late 1990s (DNBC) and to early 2000s (MoBa). The cohorts range in size from 10,625 (TIHS) to 110,000 (MoBa) mother-child pairs. Time points for contacting mothers varied, with whole cohort follow-up ending for the TIHS cohort at 12 weeks, at 7 years for the CPP and ongoing for ALSPAC, DNBC and MoBa (Table 1a).

Additional emerging cohorts.

Preliminary information about the targeted sample size, planned recruitment years, timing and source of recruitment, and data collection points for the new cohorts are in Table 1b. In summary, nine new cohorts within seven countries are collecting data on 627,500 mother-child pairs, with six recruiting mothers during pregnancy and the remaining cohorts at birth (ELFE from Group A and Gen V, TBCS from Group B).

Childhood cancer ascertainment by major category.

The 675 CC cases ascertained in the six participating cohorts to date (see Table 2) include 198 leukemias (141 acute lymphoblastic leukemia), 65 lymphomas, 161 brain tumors and 251 cancers of other types. Based on the I4C target of 1 million mothers and children pooled from the participating and emerging cohorts, it is estimated that the I4C has the potential to accrue 2952 cases of CC (diagnosed <20 years) of which 791 will be CL.[25]

Information at the IDCC according to data domain and specific exposures.

Available data in the key exposure domains for mothers, fathers, and offspring is shown in Tables 3a-c. Appendix A also lists information on known and suspected risk factors for CC, the likely /possible time window of effect and whether data are currently available at the IDCC or has been collected by the cohorts but have not to date been made available to the IDCC.

Data harmonization and descriptive results.

To date, harmonized data includes over 20 ‘core’ variables. Tables 4a-c reveal variability in characteristics of subjects based on data collected within and across cohorts that may reflect secular changes in pregnancy and birth characteristics and societal changes over the six decades of recruitment. Substantial differences are apparent for mean age of mothers at birth of the index child (24.3, youngest age (CPP) to 30.5, oldest age (DNBC)); mean height (160.9 (CPP) to 168.1 cm (MoBa)); prevalence of smoking during pregnancy (11% (MoBa) to 51% (TIHS)). For offspring, the gender of the offspring enrolled in the cohort ranged from 50% (MoBa) to 69% male (TIHS- due to selection criteria favoring males given their higher risk of SIDS, the disease of focus when the cohort was established); caesarean section delivery (5% (CPP and JPS) to 21% (TIHS)); mean birthweight in grams (3108 (TIHS) to

3560 (DNBC)); history of any breast feeding to 6 months (63% (DNBC, TIHS) to 77% (MoBa)); and paid childcare during the first 6 months (0.1% (ALSPAC) to 6% (DNBC)).

As harmonization proceeded, emerging cohorts requested information about data collection strategies and forms to facilitate future pooling of data. In response, the IDCC has developed a "*New Cohort Protocol Support Package (NCPS)"* to provide researchers with a standardized format for the collection of exposure data for etiologic studies (see Appendix H).

Publications.

The first I4C publication using a pooled dataset examined the association between birthweight and risk of CC and maternal adiposity measures as potential effect modifiers. A linear relationship was demonstrated for increasing risk of total CC and childhood leukemia with each kilogram increase in birthweight adjusted for gender and gestational age. No significant interactions were seen with maternal pre-pregnancy overweight or pregnancy weight gain. Birthweight >4000g was linked with non- leukemia cancers but, only among children diagnosed at age three or older.[4]

I4C members have described a new optimized method for extracting DNA from neonatal dried blood spots for application in methylome profiling[26,27] using samples from several of the contributing cohorts. A review paper describes the characteristics of the epigenome as a key component of fetal exposure in evaluating in utero exposures and childhood cancer risk. [28] More recently, I4C members have begun cataloguing –*omics* signatures of early-life factors that could be associated with CC.[29,30] These signatures will be analyzed across the different I4C cohorts with available biological samples. This work will complement the I4C questionnaire-based epidemiological investigations and may provide mechanistic insights into CC etiology.

Ongoing data analyses.

Current efforts are focused on: examining prospectively, the association of birth order and CL and the potential modifying roles of paternal age and birthweight; parental occupational exposure to pesticides, animals, and organic dust and risk of CC utilizing geocoded residential addresses (using DNBC for first analysis) to evaluate pesticide use near the residences during the pregnancy as well as parental occupational exposure; prenatal maternal folic acid supplementation and risk of CC; maternal infections during pregnancy and CC; epigenetic precursors of CL.

Process for requesting data for new research proposals

The I4C Steering Committee facilitates data sharing provided that all approvals are in place. The process for requesting data from any of the I4C contributing cohorts and the parallel steps undertaken at the IDCC to provide the data are in Appendix B.

COMMENT

The I4C is a valuable resource comprising both questionnaire-based epidemiological data and biological samples offering unique opportunities to advance our understanding of the etiology and mechanisms of carcinogenesis in children. It is the first established pregnancy/birth cohort consortium to have published findings on CC using harmonized variables across six cohorts.

The six participating cohorts provide an extensive set of covariates that can be leveraged with different follow-up periods ranging from pregnancy to adolescence. Ongoing collaborative work involves molecular cancer epidemiology studies and the potential for evaluation of other biomarkers.

One of the aims of the I4C has been to verify the associations reported by case-control studies for the more commonly examined exposures such as birthweight. Our analysis of birthweight included 377 cases of any cancer (115 CL and 98 ALL) and showed a linear relationship for each kilogram increment for any leukemia (Hazard ratio [HR]=1.35; 95%CI 0.90, 2.02) with similar trends observed for ALL.[4] Risk estimates from our study of birthweight were similar to those reported in the pooled analyses from the Childhood Leukemia International Consortium (CLIC) (7348 cases of CL and 12,489 controls) with an odds ratio (OR) of 1.24 for large-for-gestational age children and from a second pooled analysis from the USA, UK and Germany (4,075 cases and 12,065 controls) with an OR of 1.2 per 1000g increase in birthweight, [31,32] although a UK and US registry-based case-control study (40,000 cases and 87,000 controls) reported lower increases of CL per 500g increases of OR= 1.10 for US and 1.07 for UK data.[33]

There is a critical role for prospective assessment of exposure using pre-diagnostic questionnaire data and biological samples, but the rarity of CC and identification of an expanding number of molecularly different CC subtypes underscores the strengths and limitations of the I4C. Pooling of multiple pregnancy and birth cohorts offers prospectively collected risk factor and mechanistic data to that obtained from case-control studies. For example, information about maternal diet, viral infections, and use of folic acid and other vitamin supplements periconceptionally or during pregnancy may not be accurately recalled or available in medical records and thus not captured well in case-control studies. Relatively minor infections during infancy, details of breast-feeding, and daycare may similarly not be accurately recalled years later. Despite these potential strengths, cohort studies may also suffer from methodologic shortcomings including selection bias (cohort members are generally volunteers), under-ascertainment or misclassification of cancer outcomes, loss to follow-up over time, limited time points of data collection and measurement error (depending on the exposure assessment methods and follow-up time periods). By jointly undertaking projects with investigators leading case-control studies, the strengths of each study design can be maximized and the limitations and potential biases can be identified and quantified.

Future Directions

The I4C includes a growing number of participating cohorts and is poised to significantly increase its sample size within the next five years. I4C studies are incorporating a growing range of exposure assessment methods and tools, including Geographic Information Systems (GIS) to assess agricultural and pesticide exposures near residences, satellite measurements to measure ambient ultraviolet radiation and assignment of occupational exposures using job exposure matrices. Statistical approaches include sophisticated methods for quantifying temporal and age effects in the assessment of associations between exposure and outcome. Collaborative efforts have recently been undertaken to develop joint projects with the Childhood Leukemia International Consortium during future planned joint meetings. The prospects for combining multiple sources of pre-diagnostic exposure data and biological samples in conjunction with collaboration with other birth cohort and pediatric cancer case-control consortia offer the potential for future break-throughs in pediatric cancer etiologic research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

This work was supported by: the NIH intramural research program (NCI, NICHD) –USA; National Children's Study – USA; Tour de Cure – Australia; the Children's Cancer Centre Foundation – Australia; Bluey Day Foundation – Australia; Baxter Family Foundation –Australia; The Rotary Club of North Brighton –Australia; Private philanthropic donations – Australia; and Murdoch Children's Research Institution (M1300049), Australia. The UK Medical Research Council and the Wellcome Trust (Grant ref: 092731) and the University of Bristol provide core support for ALSPAC. The Maria Ascoli Foundation, Jerusalem, Israel, provided support for data pooling of the JPS. This work was partly supported by the Research Council of Norway through its Centres of Excellence funding scheme, project number 262700, and Innovation Fund Denmark (grant no. 09-067124). The work within the I4C carried out by the Epigenetics Group at IARC is supported by a grant from the Institut National du Cancer (INCa, Plan Cancer-EVA-INSERM, France) to Z.H. and A.G.

REFERENCES

1. Ward E, DeSantis C, Robbins A, Kohler B, Jemal A. Childhood and adolescent cancer statistics, 2014. CA Cancer J Clin 2014;64:83–103. [PubMed: 24488779]
2. Roman E, Lightfoot T, Picton S, Kinsey S. Childhood cancers. In: Thun M, Linet M, Cerhan J, Haiman C & Schottenfeld D , ed. Schottenfeld and Fraumeni Cancer Epidemiology and Prevention, Fourth Edition. New York: Oxford University Press; 2018:1119–1154.
3. Postema FAM, Hopman SMJ, Aalfs CM, Berger LPV, Bleeker FE, Dommering CJ, et al. Childhood tumours with a high probability of being part of a tumour predisposition syndrome; reason for referral for genetic consultation. Eur J Cancer 2017;80:48–54. [PubMed: 28544908]
4. Paltiel O, Tikellis G, Linet M, Golding J, Lemeshow S, Phillips G, et al. Birthweight and Childhood Cancer: Preliminary Findings from the International Childhood Cancer Cohort Consortium (I4C). Paediatr Perinat Epidemiol 2015;29:335–345. [PubMed: 25989709]
5. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 2000;343:78–85. [PubMed: 10891514]
6. Bailey HD, Fritschi L, Metayer C, Infante-Rivard C, Magnani C, Petridou E, et al. Parental occupational paint exposure and risk of childhood leukemia in the offspring: findings from the Childhood Leukemia International Consortium. Cancer Causes Control 2014;25:1351–1367. [PubMed: 25088805]

7. Bailey HD, Metayer C, Milne E, Petridou ET, Infante-Rivard C, Spector LG, et al. Home paint exposures and risk of childhood acute lymphoblastic leukemia: findings from the Childhood Leukemia International Consortium. Cancer Causes Control 2015;26:1257–1270. [PubMed: 26134047]
8. Marcotte EL, Thomopoulos TP, Infante-Rivard C, Clavel J, Petridou ET, Schuz J, et al. Caesarean delivery and risk of childhood leukaemia: a pooled analysis from the Childhood Leukemia International Consortium (CLIC). Lancet Haematol 2016;3:e176–185. [PubMed: 27063976]
9. Rudant J, Lightfoot T, Urayama KY, Petridou E, Dockerty JD, Magnani C, et al. Childhood acute lymphoblastic leukemia and indicators of early immune stimulation: a Childhood Leukemia International Consortium study. Am J Epidemiol 2015;181:549–562. [PubMed: 25731888]
10. Metayer C, Milne E, Dockerty JD, Clavel J, Pombo-de-Oliveira MS, Wesseling C, et al. Maternal supplementation with folic acid and other vitamins and risk of leukemia in offspring: a Childhood Leukemia International Consortium study. Epidemiology 2014;25:811–822. [PubMed: 25207954]
11. Olshan AF, Anderson L, Roman E, Fear N, Wolff M, Whyatt R, et al. Workshop to identify critical windows of exposure for children's health: cancer work group summary. Environ Health Perspect 2000;108 Suppl 3:595–597.
12. Hatch EE, Kleinerman RA, Linet MS, Tarone RE, Kaune WT, Auvinen A, et al. Do Confounding or Selection Factors of Residential Wiring Codes and Magnetic Fields Distort Findings of Electromagnetic Fields Studies? Epidemiology 2000;11:189–198. [PubMed: 11021619]
13. Infante-Rivard C, Jacques L. Empirical study of parental recall bias. Am J Epidemiol 2000;152:480–486. [PubMed: 10981463]
14. Searles Nielsen S, Mueller BA, De Roos AJ, Viernes HM, Farin FM, Checkoway H. Risk of brain tumors in children and susceptibility to organophosphorus insecticides: the potential role of paraoxonase (PON1). Environ Health Perspect 2005;113:909–913. [PubMed: 16002382]
15. Zuna J, Ford AM, Peham M, Patel N, Saha V, Eckert C, et al. TEL deletion analysis supports a novel view of relapse in childhood acute lymphoblastic leukemia. Clin Cancer Res 2004;10:5355–5360. [PubMed: 15328172]
16. Brown RC, Dwyer T, Kasten C, Krotoski D, Li Z, Linet MS, et al. The international childhood cancer cohort consortium (I4C). Int J Epidemiol 2007;36:724–730. [PubMed: 17255350]
17. Klebanoff MA, Read JS, Mills JL, Shiono PH. The risk of childhood cancer after neonatal exposure to vitamin K. N Engl J Med 1993;329:905–908. [PubMed: 8361503]
18. Shiono PH, Chung CS, Myrianthopoulos NC. Preconception radiation, intrauterine diagnostic radiation, and childhood neoplasia. J Natl Cancer Inst 1980;65:681–686. [PubMed: 6932521]
19. Fritz A, Percy C, Jack A, Shanmugaratnam K, Sobin L, Parkin M, et al. International Classification of Diseases for Oncology. Third Edition ed: World Health Organization; 2000. 2000.
20. International Agency for Cancer Registries. http://www.iacr.com.fr/ (Accessed 01 July 2018).
21. Dickinson HO, Salotti JA, Birch PJ, Reid MM, Malcolm A, Parker L. How complete and accurate are cancer registrations notified by the National Health Service Central Register for England and Wales? J Epidemiol Community Health 2001;55:414–422. [PubMed: 11351000]
22. Fortier I, Burton PR, Robson PJ, Ferretti V, Little J, L'Heureux F, et al. Quality, quantity and harmony: the DataSHaPER approach to integrating data across bioclinical studies. Int J Epidemiol 2010;39:1383–1393. [PubMed: 20813861]
23. Esteve A, Sobek M. Challenges and methods of international census harmonization. Historical Methods: A Journal of Quantitative and Interdisciplinary History 2003;36:66–79.
24. Schaap LA, Peeters GM, Dennison EM, Zambon S, Nikolaus T, Sanchez-Martinez M, et al. European Project on Osteoarthritis (EPOSA): methodological challenges in harmonization of existing data from five European population-based cohorts on aging. BMC Musculoskeletal Disorders 2011;12:272. [PubMed: 22122831]
25. Cancer incidence in five countries Volume XI. at http://ci5.iarc.fr/CI5-XI/Pages/summary_table_pop_sel.aspx (Accessed 20 July 2018)
26. Ghantous A, Saffery R, Cros MP, Ponsonby AL, Hirschfeld S, Kasten C, et al. Optimized DNA extraction from neonatal dried blood spots: application in methylome profiling. BMC Biotechnol 2014;14:60. [PubMed: 24980254]

27. Ghantous A, Hernandez-Vargas H, Herceg Z. DNA Methylation Analysis from Blood Spots: Increasing Yield and Quality for Genome-Wide and Locus-Specific Methylation Analysis. Methods Mol Biol 2018;1708:605–619. [PubMed: 29224166]

28. Ghantous A, Hernandez-Vargas H, Byrnes G, Dwyer T, Herceg Z. Characterising the epigenome as a key component of the fetal exposome in evaluating in utero exposures and childhood cancer risk. Mutagenesis 2015; 30:733–742. [PubMed: 25724893]

29. Joubert BR, den Dekker HT, Felix JF, Bohlin J, Ligthart S, Beckett E, et al. Maternal plasma folate impacts differential DNA methylation in an epigenome-wide meta-analysis of newborns. Nat Commun 2016;7:10577. [PubMed: 26861414]

30. Sharp GC, Salas LA, Monnereau C, Allard C, Yousefi P, Everson TM, et al. Maternal BMI at the start of pregnancy and offspring epigenome-wide DNA methylation: findings from the pregnancy and childhood epigenetics (PACE) consortium. Hum Mol Genet 2017;26:4067–4085. [PubMed: 29016858]

31. Milne E, Greenop KR, Metayer C, Schuz J, Petridou E, Pombo-de-Oliveira MS, et al. Fetal growth and childhood acute lymphoblastic leukemia: findings from the childhood leukemia international consortium. Int J Cancer 2013;133:2968–2979. [PubMed: 23754574]

32. Roman E, Lightfoot T, Smith AG, Forman MR, Linet MS, Robison L, et al. Childhood acute lymphoblastic leukaemia and birthweight: insights from a pooled analysis of case-control data from Germany, the United Kingdom and the United States. Eur J Cancer 2013;49:1437–1447. [PubMed: 23266048]

33. O'Neill KA, Murphy MF, Bunch KJ, Puumala SE, Carozza SE, Chow EJ, et al. Infant birthweight and risk of childhood cancer: international population-based case control studies of 40 000 cases. Int J Epidemiol 2015;44:153–168. [PubMed: 25626438]

Table 1a.

Descriptive characteristics of cohorts currently contributing to the I4C data pool.

Country/ Recruiting years	Study title (Acronym)	Cohort description	Pre-natal follow-up time points *(weeks gestation)*	Age at Post-natal follow-up time points	Method/s of follow-up
Australia * 1988-1995	Tasmanian Infant Health Study (TIHS) [20]	■ Birth cohort ■ **10,625** live births ■ Recruited from 6 obstetric hospitals in the state of Tasmania, Australia ■ Eligibility: Infants born from January 1, 1988 to October 31 1992, resident in Tasmania, not placed for adoption and assessed as having a greater risk associated with Sudden Infant Death Syndrome (SIDS) as determined by a predictive score that included maternal age, male gender, birthweight and season of birth (March-April, May-June, August-February), duration of second stage of labor and intention to breast feed ■ Cohort represents a one in five sample of all births considered at higher risk of SIDS in Tasmania, Australia ■ Study staff conducted interviews four days post-natal in the hospital, at home and via telephone.	n/a	■ 4 weeks ■ 10-12 weeks	■ Home visits ■ Phone interviews ■ Record linkage
Denmark * 1996-2002	Danish National Birth Cohort (DNBC) [17]	■ Population-based cohort ■ **101,042** births (I4C sample =9,596≠) ■ Sample recruited from all pregnant women across 16 counties in Denmark at first visit to the GP ■ Eligibility: All pregnant women who at first GP visit planned to carry pregnancy to term and who spoke Danish sufficiently well to complete telephone interviews were invited to participate.	■ 12 ■ 30 ■ Birth	■ 6 months ■ 18 months ■ 7 years	■ CATI** ■ Linkage to health registries
Israel 1964-1976	Jerusalem Perinatal Study (JPS) [18]	■ Population-based cohort ■ **92,408** live- and stillbirths and their parents ■ Eligibility criteria: all births to Israeli women residents of West Jerusalem ■ Women identified from compulsory notifications to District Health Office. ■ Data obtained from antenatal interviews, pediatric admissions to hospital and postpartum interviews.	For 1965-68 sub-cohort, 13,500 mothers 4th-5th month of pregnancy	■ Hospitalization ages 0-5 for sub-cohort ■ Military recruitment office age 17 for recruits.	■ Linkage with disease registries, death registries; hospitalization records ■ Military recruitment data
Norway * 1999-2008	Norwegian Mother and Child Cohort Study (MoBa)[19]	■ Population-based pregnancy cohort ■ **109,981** live-born infants (I4C sample =11,218 ≠) ■ Women at participating clinics across Norway, invited to participate in study based on request for ultrasound examination made by doctor or women recorded at Medical Birth Registry ■ Eligibility criteria: Voluntary participation at 17 weeks gestation.	■ 17 ■ 30 ■ Birth	■ 6 months ■ 18 months ■ 36 months ■ 5 years ■ 7-8 years ■ 13 years	■ Self-administered questionnaires ■ Linkages to health registries
UK * 1990-1992	Avon Longitudinal Study of Parents and Children (ALSPAC) [14,15]	■ Multi-generational birth cohort ■ 14,541 pregnancies and **14,062** live-birth ■ Representative sample of pregnancy women resident in a geographical region of Avon in South West England ■ Eligibility based on expected date of delivery between 1st April 1991 and 31st December 1992 ■ Opportunistic recruitment approach. Media campaigns, visits to community locations, antenatal maternity health services promoted study and distributed 'Expression of interest' card encouraging women to contact the study.	■ 8 ■ 12 ■ 18 ■ 32 ■ Birth	■ 4, 8 weeks ■ 6, 8,15, 18, 21, 24 months ■ 3,4,5… to 15 years	■ Self-administered questionnaires ■ Clinical assessment ■ Record linkage

Country/ Recruiting years	Study title (Acronym)	Cohort description	Pre-natal follow-up time points *(weeks gestation)*	Age at Post-natal follow-up time points	Method/s of follow-up
USA * 1959-1965	Collaborative Perinatal Project (CPP)[16]	■ Prospective, longitudinal study of prenatal women ■ 48,000 women and **60,000** births ■ Women enrolled from14 obstetrical and pediatric departments across twelve universities across the USA ■ Eligibility criteria: attended first antenatal visit at one of the participating departments and deemed eligible by the obstetrician invited to participate	n/a	■ 4, 8 months ■ 12, 36, 48 months ■ 5,6,7 years	■ Administered interviews by trained personnel
TOTAL MOTHER-CHILD PAIRS		**388,118**			

*Cohorts with prospectively collected biological samples

**CATI= Computerized telephone interviews

≠Numbers represent the data currently available at the IDCC on a random 10% sample of the non-cancer cohort plus all the incident cases of childhood cancer.

Table 1b.

Descriptive characteristics of additional emerging cohorts with potential to contribute to the I4C data pool.

Country/ Recruiting years	Study title (Acronym)	Sample size	Recruitment period	Pre-natal follow-up time points *(weeks gestation)*	Age at Post-natal follow-up time points	Method/s of follow-up	Method/s of pediatric cancer ascertainment
GROUP A: Cohorts that are well established, collecting relevant data/biospecimens, able to ascertain cases of CC and are positioned to begin contributing data to the I4C pool within the next few years.							
China * 2012-2017	Born in Guangzhou Cohort Study (BIGCS)	30,000	Early pregnancy	■ 20 ■ 24-28 ■ ≥ 33 ■ Birth	■ 6 weeks ■ 6,12,36 months ■ 3, 6, 10, 12-13, 15-16, 18 years	■ Self-administered questionnaires ■ Clinical assessments	Guangzhou Cancer Registry
France * 2011	Etude Longitudinale Française depuis l'enfance (ELFE)	18,300	Birth	n/a	■ 2 months ■ 1,2,3,5 years	■ Face-to-face and phone interviews ■ Clinical assessments	National Childhood Cancer Registry and National Health Insurance
Italy * 2005-2016	Nascita ed Infanzia: gli Effetti dell'Ambiente (NINFEA)	7,500	Any time during pregnancy	n/a	■ 6 months ■ 18 months ■ 4, 7, 10 years (planned at 13, 16 years)	On-line self-administered questionnaires	Population-based regional cancer registry and nationwide clinical registry
Japan * 2011-2014	Japan Environment and Children's Study (JECS)	100,000	Early pregnancy	■ First trimester ■ 2nd- 3rd trimester ■ Birth	■ 1, 6, 12 months ■ Every 6 months to age 13 years	Self-administered questionnaires	Physician completed questionnaire based on WHO cancer classification
Korea * 2015 -2019	Korean Children's Environmental Health Study (Ko-CHENS)	70,000	Early pregnancy	■ First trimester ■ 3rd trimester	■ 6 months ■ 1 year ■ 2 years ■ 3 years ■ up to 18 years	Self-administered questionnaires	National Health Insurance Service
	Sub-total	**225,800**					
GROUP B: Cohorts currently in various stages of development or early recruitment and follow-up, are collecting or plan to collect relevant data/biospecimens and identify cases of CC that have the potential to contribute data to the I4C pool in the future.							
Australia * Planned 2020 - 2022	Generation Victoria (GenV)	Estimated 160,000	Birth	■ Trimester 1, 2, 3	■ To be determined	■ Administrative data linkage ■ Clinical data records	Proposed linkage to state and national cancer registries
Brazil * 2014+	Campinas Infant Health Study	100,000	Early pregnancy	■ 12 ■ 30 ■ Birth	■ 6 months ■ 18 months … to 18 years	Electronic, self-administered questionnaires	Linkage to Primary Care Units
China * 2010 - 2013	Wuhan Study	120,000	Early pregnancy	■ <12 weeks	■6 weeks	■ Questionnaires administered during home visits ■ Clinical examinations ■ Electronic health	Wuhan Cancer Registry

Country/ Recruiting years	Study title (Acronym)	Sample size	Recruitment period	Pre-natal follow-up time points *(weeks gestation)*	Age at Post-natal follow-up time points	Method/s of follow-up	Method/s of pediatric cancer ascertainment
						records ■ Monitoring	
Taiwan 2013 - 2010	Taiwan Birth Cohort Study (TBCS)	20,000	Birth	n/a	■ 6 months ■ 18 months ■ 3 years ■ 5 years	■ Self-reported questionnaires	To be determined
	Sub-total	**400,000**					
	TOTAL	**627,500**					

*Cohorts with prospectively collected biological samples

Table 2.

Summary of pediatric cancers identified in the six I4C cohorts currently contributing to the I4C data pool.

	ALSPAC (UK)	CPP (USA)	DNBC (Denmark)	JPS (Israel)	MoBa (Norway)	TIHS (Australia)	POOLED TOTAL
Number of childhood cancer cases	24	50	202	172	200	27	**675**
Number of Leukemia cases	5	16	64	39	70	4	**198**
- ALL	*n/a*	*16*	*46*	*26*	*52*	*3*	***141***
Number of Lymphoma cases	n/a	5	19	6	31	4	**65**
Number of Brain/CNS tumors	n/a	2	62	56	35	6	**161**
Number of other cancers	19	27	57	71	64	13	**251**
Method/s of pediatric cancer ascertainment	Linkage to National Cancer Registry	- Examination of diagnostic summaries - Review of death records	Linkage to National Cancer Registry	Linkage to National Cancer Registry	Linkage to National Cancer Registry	Linkage to State Cancer Registry	--

n/a : Not available due to data protection/IRB issues associated with small numbers.

Table 3a:

Maternal-related domains associated with the international question set for examining the etiology of childhood cancers.

	ALSPAC	CPP	DNBC	JPS	MoBa	TIHS
MATERNAL						
Age *- at birth of index child*	Y	Y	Y	Y	Y	Y
Education	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
- years of education	N	Y	Y	Y	Y	Y
Marital status during pregnancy	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
- married/co-habitating/common law	Y *	Y *	Y	Y *	Y	Y *
- divorced/separated	Y	Y	N	Y	Y	Y
- widow	Y	Y	N	Y	Y	Y
- single	Y	Y	N	Y	Y	Y
Income (gross per household)	Y	Y	Y	N	Y	Y
Anthropometric measures	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
- height	Y	Y	Y	Y	Y	Y
- weight before/ in early pregnancy	Y	Y	Y	Y	Y	Y
- weight at delivery	Y	Y	Y	Y	Y	Y
Smoking	**Y**	**Y**	**Y**	**Y (limited)**	**Y**	**Y**
- ever smoked	Y	Y	N	Y	Y	N
- smoked just prior to conception	Y	N	N	N	Y	N
- smoked during pregnancy (weeks gestation)	Y	N	Y	Y	Y	Y
- number of cigarettes per day	Y	Y	Y	Y	Y	Y
- age started/stopped smoking	Y	Y	Y	N	Y	N
- cigarette brand	Y	N	Y	N	Y	N
- smoking post-natal	Y	N	Y	N	Y	Y
Passive smoking during pregnancy	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
- hours per week exposed	Y	N	Y	N	N	Y
Illicit drug use	**Y**	**N**	**Y**	**N**	**Y**	**N**
- weeks gestation	Y		Y		Y (limited)	
- type of drug	Y		Y		Y	

	ALSPAC	CPP	DNBC	JPS	MoBa	TIHS
Alcohol consumption during pregnancy	**Y**	**N**	**Y**	**N**	**Y**	**Y**
- weeks gestation	Y		Y		Y	Y
- frequency of consumption	Y		Y		Y	Y
- type of alcohol (beer, wine, spirits)	Y		Y		Y	N
- units of alcohol consumed	Y		Y		Y	N
- binge drinking	Y		Y		Y	Y
Diet during pregnancy	**Y**	**N**	**Y**	**N**	**Y**	**Y**
-weeks gestation	Y		Y		Y	
Vitamin supplement use during pregnancy	**Y**	**N**	**Y**	**N**	**Y**	**Y**
Folic acid	Y		Y		Y	Y
multivitamins	Y		Y		Y	Y
- weeks gestation when used	Y		Y		Y	Y
Other medical conditions during pregnancy	**Y**	**Y**	**Y**	**Y**	**Y**	**N**
- diabetes	Y	Y	Y	Y	Y	
- hypertension	Y	Y	Y	Y	Y	
- pre-eclampsia	Y	Y	Y	Y	Y	
Atopy/Asthma in mother	Y	Y	Y	Y	Y	N
Prescription medications	**Y**	**Y**	**Y**	**Y (limited)**	**Y**	**Y**
- name	Y	Y	Y		Y	Y
- condition used for	Y	N	Y		N	N
- weeks gestation	Y	Y	Y		Y	Y
Reproductive history	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
- number of prior pregnancies	Y	Y	Y	Y	Y	Y
- previous losses (abortions, miscarriages, stillborn)	Y	Y	Y	Y	Y	N
-gestation age at time of loss	N	N	Y	Y	Y	N
Occupation	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
- job title	Y	Y	Y	Y	Y	Y
- industry	Y	N	Y	Y	Y	Y
- occupation classification code	1990 Standard Occupational Classification	List of occupations	Standard Occupational Classification	List of occupations	ISCO88	Australian Standard Classification of Occupations
- during pregnancy	Y	Y	Y	Y	Y	Y

	ALSPAC	CPP	DNBC	JPS	MoBa	TIHS
- employment status during pregnancy	Y	N	Y	N	Y	Y
- post-natal occupation	Y	N	Y	N	Y	Y
- job history	Y	N	N	N	Y	N
Infections during pregnancy	**Y**	**Y**	**Y**	**Y (limited)**	**Y**	**Y (limited)**
- types	Y	Y	Y		Y	Y
- weeks gestation	Y	N	Y		Y	Y
-use of medication	Y	Y	Y		Y	Y
Pesticide/chemical exposures	**Y**	**N**	**Y**	**N**	**Y**	**N**
- weeks gestation	Y		Occupation		Y	
- frequency of use	Y		N		Y	
Antenatal radiation exposure	Y	Y	N	Y	Y	Y
Sun exposure/Vitamin D intake	N	N	N	N	N	N
Biological specimens- during pregnancy	Y	Y	Y	N	Y	Y

Y= Yes, *Epidemiological data for exposure domain or biological specimen collected by respective cohort.*

N= No, *Epidemiological data for exposure domain or biological specimen not collected or available in respective cohort.*

* Marital status includes marriage but not cohabitating or common law.

Table 3b:

Paternal-related domains associated with the international question set for examining the etiology of childhood cancers.

PATERNAL	ALSPAC	CPP	DNBC	JPS	MoBa	TIHS
Age	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
- at birth of index child	Y	Y	Y	Y	Y	Y
Education						
- years of education	Y	Y	Y	Y	Y	Y
Smoking	**Y**	**N**	**Y**	**Y** (for sub-cohort)	**Y**	**Y**
- smoked before pregnancy	Y		N		Y	N
- smoked during pregnancy	Y		Y		Y	Y
- number of cigarettes per day	Y		N		Y	Y
- smoked around child	Y		**?**		Y	Y
History of diabetes	Y	Y	N	N	Y	Y (family level)
Atopy/Asthma in father	Y	N	Y	N	Y	Y
Occupation	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
- job title	Y	Y	Y	Y	Y	**Y**
- industry	Y	**N**	Y	Y	Y	N
- occupation classification code	1990 Standard Occupational Classification		4-figure occupation code	-	4-figure occupation code	4-figure occupation code
- job history	Y	N	N	N	N	N
- employment status	Y	Y	Y	Y	Y	Y
Biological specimens	Y (collected post-natal)	N	Y	N	Y	N

Y= Yes, *Epidemiological data for exposure domain or biological specimen collected by respective cohort.*

N= No, *Epidemiological data for exposure domain or biological specimen not collected or available in respective cohort.*

Table 3c:

Birth and child-related domains associated with the international question set for examining the etiology of childhood cancers.

BIRTH-RELATED	**ALSPAC**	**CPP**	**DNBC**	**JPS**	**MoBa**	**TIHS**
Gestational age	Y	Y	Y	Y (limited)	Y	Y
Gender	Y	Y	Y	Y	Y	Y
Weight	Y	Y	Y	Y	Y	Y
Length	Y	Y	Y	N	Y	Y
Biological specimens	Y	Y	Y	N	Y	Y
CHILD-RELATED						
Growth measures	**Y**	**Y**	**Y**	**N**	**Y**	**Y** (limited)
- length	Y	Y	Y		Y	Y
- weight	Y	Y	Y		Y	Y
- head circumference	Y	Y	Y		Y	Y
- skin fold measures	Y	N	N		N	Y
Infant feeding habits	**Y**	**Y**	**Y**	**N** (for previous pregnancies on sub-cohort only)	**Y**	**Y**
- breast	Y	Limited	Y		Y	Y
- formula	Y	N	Y		Y	Y
- other	Y	N	Y		Y	Y
-duration	Y	N	Y		Y	Y
Passive smoking exposure	Y	N	Y	Y	Y	Y
Day care attendance	**Y**	**N**	**Y**	**N**	**Y**	**Y**
- age at commencement	Y		Y		Y	N
- length of attendance	Y		Y		Y	N
Infections during first year of life	**Y**	**Y**	**?**	**Y** (limited) (only if hospitalized; for sub-cohort only)	**Y**	**Y**
- type of infection	Y	Y			Y	Y
- age	Y	N			Y	Y
- use of medication for infection	Y	N			Y	Y
Pesticide/chemical exposure	Y	N	Y	N	Y	Y
Infant diet	**Y**	**N**	**Y**	**N**	**N**	**N**

BIRTH-RELATED	ALSPAC	CPP	DNBC	JPS	MoBa	TIHS
-age	Y		Y			
Sun exposure	Y	N	N	N	N	N
Radiation exposure during first year of life	Y (x-ray of hips at 6 months)	N	Y	N	N	N
Asthma/Eczema	Y	Y	Y	N	Y	Y

Y= Yes, *Epidemiological data for exposure domain or biological specimen collected by respective cohort.*

N= No, *Epidemiological data for exposure domain or biological specimen not collected or available in respective cohort.*

Table 4a:

Key maternal-related, harmonized characteristics from cohorts contributing to the current pooled dataset.

	ALSPAC	CCP	DNBC	JPS	MoBa	TIHS
Full cohort population [Total =374,600]	**14,049**	**53, 679**	**94,690**	**90,079**	**111,399**	**10,624**
Recruitment years	**1990 – 1992**	**1959 – 1966**	**1996 – 2002**	**1964 – 1976**	**1999 – 2007**	**1988 – 1995**
Number of mother-child pairs with available data in pooled dataset [Total=189,326]	**14,042** *	**53,677**	**9,588**	**90,079**	**11,309**	**10,624**
Person-years of follow-up ** **[Total =2,880,400]**	**112,341**	**350,027**	**756,848**	**720,169**	**856,096**	**84,928**
Age at time of index child's birth, years; *mean ± SD*	28.0 ± 5.0	24.1 ± 5.9	30.5 ± 4.3	27.7 ± 5.7	30.3 ± 4.6	24.3 ± 4.9
Education: completed 12 or more years of education; *% (n)*	---	41 (21,700)	65 (4,391)	42 (37,423)	80 (8,122)	21 (2,177)
Married/cohabitating; *% (n)*	70 (9,852) [married]	77 (41,410)	94 (8,285)	99 (88,516)	92 (10,259)	80 (8492)
Height, cm; *mean ± SD*	164.0 ± 6.7	160.9 ± 6.9	168.8 ± 6.0	161.9 ± 6.0	168.1 ±5.9	162.3 ±7.2
Pre-pregnancy BMI, kg/m^2; *mean ± SD*	22.9 ± 3.8	22.7 ± 4.3	23.6 ± 4.4	22.1 ± 3.1	24.0 ± 4.2	23.3 ± 4.8
Total pregnancy weight change, kg; *mean ± SD*	12.5 ± 4.8	9.7 ± 5.1	15.0 ± 6.1	11.2 ± 4.4	15.0 ± 6.0	14.2 ± 6.7
Any smoking during pregnancy; *% (n)*	30 (3,670)	46 (34,317)	26 (2,413)	12 (3,389) sub-cohort of 20,000	11 (973)	51 (5,431)
Partner or person living in home smoked during pregnancy; *% (n)*	46 (5,512)	---	66 (5,743)	44 (7,253)	9 (842)	54 (5,722)
Any alcohol consumption during pregnancy; *% (n)*	73 (9,083)	---	56 (5,287)	---	31 (2,793)	34 (3,590)
Intake of vitamin supplement containing folic acid during pregnancy; *% (n)*	23 (2,707)	---	79 (7,613)	---	80 (9,015)	48 (5,035)
Diabetes: *% (n)*						---
- Any type	0.9 (105)	0.8 (404)	2 (176)	0.7 (627)	1.3 (151)	
- Gestational	0.5 (57)	0.2 (127)	2 (147)	0.5 (455)	0.8 (92)	
Previous miscarriage *% (n)*	22 (2,812)	18 (9,476)	18 (1,753)	20 (9,064)	11 (1,216)	---
Mother had x-ray of hip, pelvis, abdomen, or back area	2 (210)	20 (10,550)	---	0.1 (16)	0.4 (44)	---
Parity *** *% (n)*						
0	45 (5,763)	4 (2,192)	47 (4,271)	29 (25,750)	45 (5,099)	45 (4749)
1	35 (4,476)	23 (12,161)	36 (3,295)	24 (21,605)	36 (4,059)	33 (3454)
2	14 (1,818)	16 (8,559)	13 (1,196)	17 (15,282)	16 (1,753)	15 (1595)
3	4 (529)	11 (5,909)	3 (234)	10 (9,124)	3 (317)	5 (541)
4+	1.8 (227)	19 (10,114)	0.6 (55)	20 (17,689)	0.6 (71)	3 (283)

Numbers reported as % (n) for categorical variables and mean ± sd for continuous variables based on total number of data points available for each variable.

* Difference between numbers in full ALSPAC cohort and data available for I4C is due to numbers in I4C not including triplets and quadruplets due to confidentiality reasons.

** Person-years of follow-up is based on start time defined as birth date (date is set to zero years); end time for those with cancer defined as the date of cancer diagnosis; end time for those without cancer defined as the date the child is no longer under observation.

*** Parity defined as the number of previous pregnancies resulting in a live or stillbirth

"---" indicates data was not collected for this particular cohort or are not reliable if collected, or not able to be harmonized with other cohorts.

Table 4b:

Paternal-related: harmonized characteristics from cohorts contributing to the current pooled dataset.

	ALSPAC	CCP	DNBC	JPS	MoBa	TIHS
Number with available data in pooled dataset	**11,720**	**39,197**	**9,436**	**90,079**	**11,246**	**10,393**
Age at time of index child's birth, years; *mean ± sd*	30.7 ± 5.7	28.3 ± 7.0	32.8 ± 5.2	31.6 ± 6.8	32.8 ± 5.3	27.1 ± 5.9
Education: completed 12 or more years of education; *% (n)*	--	46 (20,067)	45 (2,909)	47(41,264)	64(6,253)	21 (2,043)
History of diabetes; *% (n)*	0.4 (38)	0.5 (239)	---	---	---	---

Numbers reported as % (n) for categorical variables and mean ± standard deviation for continuous variables based on total number of data points available for each variable.

"---" indicates data was not collected for this particular cohort or are not reliable if collected.

Table 4c:

Birth and infant-related: harmonized characteristics from cohorts contributing to the current pooled dataset.

	ALSPAC	CCP	DNBC	JPS	MoBa	TIHS
Number with available data in pooled dataset	**14,049**	**53,679**	**9,595**	**90,079**	**11,309**	**10,624**
Gestational age, weeks; *mean ± SD*	39.4 ± 2.0	39.3 ± 3.1	39.9 ± 1.9	39.6 ± 2.3	39.3 ± 2.1	38.5 ± 2.7
Birth weight, grams*; mean ± SD*	3385 ± 573	3161 ± 543	3560 ± 591	3252 ± 541	3540 ± 617	3108 ± 769
Placenta weight, grams; *mean ± SD*	662 ± 166	437 ± 96.5	669 ± 157	---	694 ± 179	---
Birth length, cm; *mean ± SD*	50.6 ± 2.5	49.8 ± 2.8	52.2 ± 2.7	---	50.1 ± 2.7	48.5 ± 3.5
Head circumference at birth, cm; *mean ± SD*	34.7 ± 1.6	33.7 ± 1.6	35.3 ± 1.8	---	35.2 ± 1.8	33.8 ± 2.3
Delivery by caesarean-section; *% (n)*	11 (1,494)	5 (2,517)	16 (1,551)	5 (4,273)	16 (1,795)	21 (2,217)
Gender of index child, male; *% (n)*	52 (7,266)	51(25,952)	51 (4,883)	52 (46,392)	50 (5,670)	69 (7,307)
Index child is first born; *% (n)*	43 (5,613)	28 (14,899)	32 (3,112)	29 (25,750)	45 (5,099)	43 (4,603)
Singleton birth; *% (n)*	97 (13,678)	98 (50,351)	100 (9,584)	98 (87,890)	94 (10,581)	88 (9,362)
Down Syndrome; *% (n)*	0.1 (14)	0.1 (69)	0.1 (11)	0.2 (34)	0.1 (16)	0 (0)
Any breast feeding to 6 months; *% (n)*	75 (8,387)	---	63 (4,515)	--	77 (8,724)	63 (6,325)
Childcare during first 6 months; *% (n)*	0.1 (18)	2 (200)	6 (548)	---	0.2 (17)	---

*Numbers reported as % (n) for categorical variables and mean ± sd for continuous variables based on total number of data points available for each variable.

"---" indicates data was not collected for this particular cohort or are not reliable if collected