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

Tikellis G[1], Dwyer T[1,2], Paltiel O[3], Phillips GS[4], Lemeshow S[5], Golding J[6], Northstone K[6], Boyd A[6], Olsen S[7], Ghantous A[8], Herceg Z[8], Ward MH[9], Håberg SE[10], Magnus P[10], Olsen J[11], Ström M[7], Mahabir S[12], Jones RR[9], Ponsonby AL[1], Clavel J[13], Charles MA[13], Trevathan E[14], Qian ZM[15], Maule MM[16], Qiu X[17], Hong YC[18], Brandalise S[19], Roman E[20], Wake M[1], He JR[17,21], Linet MS[22]; International Childhood Cancer Cohort Consortium.

Abstract

Background

Childhood cancer is a rare but leading cause of morbidity and mortality. Established risk factors, accounting for <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 utilise prospective, pre-diagnostic exposure assessments and biological samples.

Methods

Eligibility criteria, follow-up methods and identification of paediatric cancer cases are described for cohorts currently participating or planning future participation. Also described are exposure assessments, harmonisation methods, biological samples potentially available for I4C research, the role of the I4C data and biospecimen coordinating centres 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 5 years. Harmonised data currently includes over 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 paediatric cancer using harmonised variables across six pregnancy/birth cohorts. Projected increases in sample size, expanding sources of exposure data (eg, 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 paediatric cancer aetiologic research.

1 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 <10% of childhood cancer (CC) incidence.**5** More recently, pooled case- control studies of childhood leukaemia (CL) suggest modestly increased risks associated with residential painting and pesticide use and pre- labour caesarean delivery**6**-**8**and slightly decreased risks from day care attendance, extended breast feeding, and maternal vitamin and folic acid supplement use.**9**, **10** Known and suspected risk factors for CC**2** are briefly summarised in Appendix **S1**.

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 aetiologic 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 mechanisms of carcinogenesis.**12**, **13** Biospecimens collected prospectively are an advantage of prospective pregnancy/birth cohort studies for exploring CC aetiology, although a few case- control studies have accessed archived pre- diagnostic newborn blood spots**14** or cord blood.**15**

Our objective was 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.

2 METHODS

2.1 Overview, structure and operations

The overarching goal of the I4C is to understand the aetiology and mechanistic underpinnings of CC by exploiting prospectively collected exposure and biomarker data. The I4C Steering Committee includes lead investigators from cohorts, clinicians, paediatric cancer epidemiologists, molecular epidemiologists, exposure assessment experts and funders (**https://www.mcri.edu.au/research/projects/international-childhood-cancer-cohort-consortium-i4c/i4c-consortium**). The International Data Coordinating Centre (IDCC) at the Murdoch Children's Research Institute (MCRI) in Melbourne Australia houses the cohort data, manages data transfers, harmonises variables, develops pooled datasets, 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 working groups and annual open scientific meetings attended by investigators from participating and additional emerging cohorts and other experts.

2.2 Study populations

2.2.1 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 aetiology- related hypotheses. The specific goals and original outcomes of the individual cohorts (eg, pregnancy complications and/or serious chronic childhood conditions) may vary, but critical data items include

parental and offspring demographic, life style, medical, reproductive, environmental factors and parental occupational information. Specific responsibilities of newly joining or participating I4C cohorts include sharing of data (and biospecimens- if available) for current and future proposals.

2.2.2 Currently contributing cohorts

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

2.3 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 anonymised data were requested (see Appendix **S2**).

2.4 Follow- up methods

Strategies and time points for follow- up varied (Table **1**). 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, JPS, 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 postnatal.

2.5 CC case ascertainment and classification

2.5.1 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 paediatricians.

2.5.2 Classification

To date, age at diagnosis for CC has been < 15 years, but going forward, will extend to < 20 years. Tumours 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- 0- 3 morphology codes for leukaemia included 9800- 9948, gliomas 9380- 9480 and lymphomas 9590- 9729. From this information, the IDCC used the following six groupings: any cancer, any leukaemia, acute lymphoblastic leukaemia (ALL), acute myeloid leukaemia (AML), any lymphomas, any central nervous system (CNS)/brain tumour, or other cancers. Due to small numbers and confidentiality issues, ALSPAC provided only any cancer, any leukaemia and acute lymphoblastic leukaemia. 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 United Kingdom, 2001 reports showed 94% coverage of cancers ascertained during 1971- 1989.**21** Since CPP cases were identified through indirect methods, some cancer cases may have been missed.

2.6 Exposure data

2.6.1 Identification of data domains and specific variables associated with CC

Thirty domains were established for key exposures (eg, birthweight, folic acid supplements, and others; see Tables **4**- 6). The IDCC will submit requests to obtain additional data if needed for future proposals (See Appendix **S3** for details of the process). While the main domains centre

around the mother and child (see Tables **4**, **6**), some information on fathers is also available (see Table **5**).

2.6.2 Harmonisation of exposure data

Our approach is similar to other consortia.**22**-**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 harmonisation was not possible. The individual cohorts collected data in a standardised, structured approach from self- reported, telephone interview or in- person administered questionnaires. Each cohort provided anonymised, individual- level. Data harmonisation was carried out centrally by the IDCC project director (GT) with assistance from senior epidemiologists (TD, ALP). Each exposure variable was harmonised individually and the data evaluated for consistency within and across variables (see Appendix **S4**).

2.7 Biological samples

Four of the participating cohorts (ALSPAC, DNBC, MoBa and 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 the following: whole blood, serum, urine and placentas from mothers; cord blood, blood (neonatal blood spots), hair, nails and teeth from the offspring (Appendix S5). All additional emerging cohorts are collecting a variety of biological samples.

2.8 Identification of additional emerging cohorts

Two groups of emerging cohorts are currently involved in I4C activities but are not as yet contributing cancer cases, exposures or biospecimens to the pool. These are detailed in Table 2. 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- Korea). Group B consists of four cohorts in various stages of development or early recruitment and follow- up from Australia, Brazil, China and Taiwan.

2.9 Housing of data at the IDCC: platform, confidentiality, privacy and security measures

The data transferred to the IDCC are securely housed on a web- based application located on the MCRI's secure e- Research portal (see Appendix S6). Access is restricted to authorised personnel following approval by the I4C Steering Committee and a representative from each study contributing to the pooled data set.

For added security, data files are encrypted before being sent to the IDCC. Most studies have excluded unique personal identifiers (eg, 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 is 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.

2.10 Statistical consultation and support on study designs, data harmonisation, 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 data set.

While complete harmonisation 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.

2.11 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 is 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 the following: (a) accounting for different cohorts; (b) handling missing data for risk factors using multiple chained imputation techniques; (c) dealing with different lengths of follow-up of the contributing cohorts; (d) examining confounding and effect-modification of postulated risk factors; (e) finding the correct scale for continuous covariates and (f) testing the proportional hazard assumption for Cox regression models. Further details and strategies are in Appendix **S7**.

3 RESULTS

3.1 Cohorts currently contributing data

Six cohorts (Table **1**) currently contribute data on 388 120 mother-child pairs as well as less extensive paternal data for certain domains (Tables **4**-**6**). Recruitment periods span over six decades from the late 1950s (CPP), mid-1960s to mid-70s (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 **1**).

3.2 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 is in Table **2**. 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 and TBCS from Group B).

3.3 Childhood cancer ascertainment by major category

The 675 CC cases ascertained in the six participating cohorts to date (see Table **3**) include 198 leukaemias (141 acute lymphoblastic leukaemia), 65 lymphomas, 161 brain tumours 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**

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

Available data in the key exposure domains for mothers, fathers and offspring are shown in Tables **4-6**. Appendix **S1** 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 have been collected by the cohorts but have not to date been made available to the IDCC.

3.5 Data harmonisation and descriptive results

To date, harmonised data includes over 20 “core” variables. Tables **7-9** 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 favouring 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 harmonisation 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 standardised format for the collection of exposure data for aetiologic studies (see Appendix **S8**).

3.6 Publications

The first I4C publication using a pooled data set 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 any CC and childhood leukaemia 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 >4000 g was linked with non-leukaemia cancers but, only among children diagnosed at age three or older.**4**
I4C members have described a new optimised 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 foetal 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 analysed 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 aetiology.

3.7 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 utilising geocoded residential addresses (using DNBC in the 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.
3.8 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 **S2**.

4 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 aetiology and mechanisms of carcinogenesis in children. It is the first established pregnancy/birth cohort consortium to have published findings on CC using harmonised 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 leukaemia (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 (4075 cases and 12 065 controls) with an OR of 1.2 per 1000 g 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 500 g 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 day care 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 maximised and the limitations and potential biases can be identified and quantified.

4.1 Future directions

The I4C includes a growing number of participating cohorts and is poised to significantly increase its sample size within the next 5 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 paediatric cancer case-control consortia offer the potential for future breakthroughs in paediatric cancer aetiologic research.

ACKNOWLEDGEMENTS

This work was supported by: the NIH intramural research programme (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). Danish cancer cases were ascertained by the Danish Childhood Cancer Registry (Steering Committee: Catherine Rechnitzer, Peter Skov Wehner, Steen Rosthøj, and Henrik Schrøder). The work within the I4C carried out by the Epigenetics Group at IARC is supported by grants from the Institut National du Cancer (INCa, Plan Cancer- EVA- INSERM, France) to Z.H. and A.G. and L'Association pour la Recherche sur le Cancer (ARC, France) to Z.H.

REFERENCES

1. 1Ward E, DeSantis C, Robbins A, Kohler B, Jemal A. Childhood and adolescent cancer statistics, 2014. *CA Cancer J Clin*. 2014; **64**: 83- 103.

2. 2Roman E, Lightfoot T, Picton S, Kinsey S. Childhood cancers. In: M Thun, M Linet, J Cerhan, C Haiman, D Schottenfeld, eds. *Schottenfeld and Fraumeni Cancer Epidemiology and Prevention*, 4th edn. New York, NY: Oxford University Press; 2018: 1119- 1154.
3. 3Postema FAM, Hopman SMJ, Aalfs CM, 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.
4. 4Paltiel O, Tikellis G, Linet M, et al. Birthweight and childhood cancer: preliminary findings from the international childhood cancer cohort consortium (I4C). *Paediatr Perinat Epidemiol.*2015; **29**: 335- 345.
5. 5Lichtenstein P, Holm NV, Verkasalo PK, 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.
6. 6Bailey HD, Fritschi L, Metayer C, 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.
7. 7Bailey HD, Metayer C, Milne E, 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.
8. 8Marcotte EL, Thomopoulos TP, Infante- Rivard C, et al. Caesarean delivery and risk of childhood leukaemia: a pooled analysis from the Childhood Leukemia International Consortium (CLIC). *Lancet Haematol*. 2016; **3**: e176- e185.
9. 9Rudant J, Lightfoot T, Urayama KY, 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.
10. 10Metayer C, Milne E, Dockerty JD, 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.
11. 11Olshan AF, Anderson L, Roman E, 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. 12Hatch EE, Kleinerman RA, Linet MS, 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.
13. 13Infante- Rivard C, Jacques L. Empirical study of parental recall bias. *Am J Epidemiol*. 2000;**152**: 480- 486.
14. 14Searles 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.
15. 15Zuna J, Ford AM, Peham M, et al. TEL deletion analysis supports a novel view of relapse in childhood acute lymphoblastic leukemia. *Clin Cancer Res*. 2004; **10**: 5355- 5360.
16. 16Brown RC, Dwyer T, Kasten C, et al. The international childhood cancer cohort consortium (I4C). *Int J Epidemiol*. 2007; **36**: 724- 730.
17. 17Klebanoff 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.
18. 18Shiono PH, Chung CS, Myrianthopoulos NC. Preconception radiation, intrauterine diagnostic radiation, and childhood neoplasia. *J Natl Cancer Inst*. 1980; **65**: 681- 686.
19. 19Fritz A, Percy C, Jack A, et al. *International Classification of Diseases for Oncology*. 3rd edn.Geneva: World Health Organization; 2000.
20. 20 International Agency for Cancer Registries. http://www.iacr.com.fr/. Accessed 01 July 2018.
21. 21Dickinson 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.
22. 22Fortier I, Burton PR, Robson PJ, et al. Quality, quantity and harmony: the DataSHaPER approach to integrating data across bioclinical studies. *Int J Epidemiol*. 2010; **39**: 1383- 1393.
23. 23Esteve A, Sobek M. Challenges and methods of international census harmonization. *J Quant Interdiscip Hist*. 2003; **36**: 66- 79.
24. 24Schaap LA, Peeters GM, Dennison EM, et al. European Project on Osteoarthritis (EPOSA): methodological challenges in harmonization of existing data from five European population- based cohorts on aging. *BMC Musculoskelet Disord*. 2011; **12**: 272.
25. 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. 26Ghantous A, Saffery R, Cros MP, et al. Optimized DNA extraction from neonatal dried blood spots: application in methylome profiling. *BMC Biotechnol*. 2014; **14**: 60.
27. 27Ghantous 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.
28. 28Ghantous 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.
29. 29Joubert BR, den Dekker HT, Felix JF, et al. Maternal plasma folate impacts differential DNA methylation in an epigenome- wide meta- analysis of newborns. *Nat Commun*. 2016; **7**: 10577.
30. 30Sharp GC, Salas LA, Monnereau C, 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.
31. 31Milne E, Greenop KR, Metayer C, et al. Fetal growth and childhood acute lymphoblastic leukemia: findings from the childhood leukemia international consortium. *Int J Cancer*. 2013;**133**: 2968- 2979.
32. 32Roman E, Lightfoot T, Smith AG, 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.
33. 33O'Neill KA, Murphy MF, Bunch KJ, 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.