MoBa and the International Childhood Cancer Cohort Consortium (I4C)

Terence Dwyer[1,2], Gabriella Tikellis[3], Akram Ghantous[2], Stanley Lemeshow[4], Siri E. Håberg[5],
Jørn Olsen[6], Ora Paltiel[7], Jean Golding[8], Martha S. Linet[9], Zdenko Herceg[2],
Monica C. Munthe-Kaas[10] and Camilla Stoltenberg[5]

1) *The George Institute for Global Health, Nuffield Department of Population Health, University of Oxford, Oxford, UK*
2) *The International Agency for Research on Cancer, Lyon, France*
3) *Murdoch Children's Research Institute, Royal Children's Hospital, University of Melbourne, Melbourne, Australia*
4) *The Ohio State University College of Public Health, Columbus, Ohio*
5) *Norwegian Institute of Public Health, Oslo, Norway*
6) *Department of Epidemiology, University of California, Los Angeles, CA, USA*
7) *School of Public Health and Department of Hematology, Hadassah-Hebrew University, Jerusalem, Israel*
8) *Centre for Child & Adolescent Health, School of Social & Community Medicine, University of Bristol, Bristol, UK*
9) *Radiation Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA*
10) *Department of Pediatrics, Oslo University Hospital, Oslo, Norway*

Correspondence: Terence Dwyer, The George Institute for Global Health, Nuffield Department for Population Health, Oxford Martin School, University of Oxford, 34 Broad Street, Oxford OX1 3BD, United Kingdom
E-mail: Terence.dwyer@georgeinstitute.ox.ac.uk   Telephone: + 44 (0) 1865 617203

CHILDHOOD CANCER

The overall incidence of childhood cancer (CC) is approximately 1 per million children per year and has been steady or increasing in most developed countries in recent decades (1-3). While there has been considerable progress in the treatment of the disease, with a much larger proportion of children surviving now than in previous decades, there has been little progress in prevention.

This is not to say that there have not been advances in gaining an understanding of at least some of the underlying biological processes. For example, it is now known that certain chromosomal translocations that can be present at birth are also found in a substantial proportion of those who develop leukaemia (4). Nonetheless, how this might relate to exposures that might be avoided is not well understood. Apart from a small fraction of CC cases that can be explained by ionizing radiation (5), compelling evidence linking environmental exposures or biological mechanisms and associated biological mechanisms to the broader occurrence of CC is lacking.

In contrast to adult cancer where greater progress has been made on prevention, investigations of putative causes in childhood cancer using epidemiology has been principally limited to case-control or record linkage approaches (6). Case-control studies often suffer from recall or selection bias, and while record linkage approaches are generally free of these, they often lack detail in exposure measures. Both case-control and record linkage studies also lack the capacity to provide the range of pre-disease biospecimens that might be of interest.

Cohort studies have the potential to fill in some of the gaps in knowledge, but they have not been used to any extent with childhood cancer because they need to be very large to have the power needed to provide useful data. The International Childhood Cancer Cohort Consortium was assembled to address this deficiency (6).

THE INTERNATIONAL CHILDHOOD CANCER COHORT CONSORTIUM (I4C)

The concept of establishing I4C was first put forward in 2004 during the planning of the National Children's Study (NCS), a large childhood cohort study in the USA intending to recruit 100,000 mothers and babies (www.nationalchildrensstudy.gov/about/overview/Pages/default.aspx). It was recognized that because of the infrequent occurrence of CC – only 250 cases of all child cancers were anticipated to occur among members of the NCS cohort up to the age of 14 which defines the period of childhood cancer occurrence (4). Importantly only 50 cases of Childhood leukaemia, the most cohesive subgroup of childhood cancers, could be anticipated in a birth cohort of 100,000. The NCS alone would not have sufficient power to detect associations with potential exposures of interest.

However, if subjects were able to be pooled with other birth cohorts then it might be possible to achieve sufficient power to investigate associations between early exposures and CC. The existence of a number of large birth cohorts globally was identified in 2004 and the total number of subjects apparently available approached the number (~500,000 mother/child pairs), that power calculations (7) indicated would be required. These very large cohorts, with around 100,000 subjects each included the Norwegian Mother and child Cohort Study (MoBa) (6), the Danish National Birth Cohort (DNBC) (8), and the Jerusalem Perinatal Study (JPS)

Table 1. Descriptive characteristics of the six I4C-member cohorts included in the pooled dataset.

	ALSPAC	CPP	DNBC	JPS	MoBa	TIHS	Total
Recruitment years	1991-1992	1959-1966	1996-2002	1964-1976	1999-2009	1987-1995	1959-2009
Number of subjects	13,664	50,342	*98.000	87,856	*105,000	9,362	*360,000
Number of subjects on whom data has been transferred	13,664	50,342	8,603	87,856	10,497	9,362	112,781
Years of follow-up, mean (range)	14.9 (0.5-15)	5.6 (0.0-8.0)	11 (8.3-14.0)	15 (15.0-15.0)	4.4 (0.5-10.1)	14.7 (12.7-15.0)	9.9 (0.0-15.0)
Number of cancer cases	24	50	151	168	127	27	547

* approximately

(6). Two additional more modest sized birth cohorts, the Tasmanian Infant Health Survey (TIHS) (9) and the Avon Longitudinal Study of Parents And Children (ALSPAC) (10), were known to the group and were also included. Each of these five studies had early life exposure measurements that were considered relevant to the study of CC, and all were able to follow subjects for cancer. We had become aware of a sixth cohort, the Chinese Children and Families Cohort Study (CFCS) of 242,000 subjects (11), but it was not clear at that time whether subjects in this cohort could be followed for CC. While it became a member of I4C it is not yet providing data to the consortium.

The I4C was formally established for the purpose of bringing together these cohorts at a meeting hosted by NCI and NICHD in 2005 at which representatives from MoBa were in attendance. At this meeting they, along with several other cohorts confirmed their intention to collaborate, through the consortium, to investigate causes of CC.

Several other cohorts were identified in the following years, including the NIH birth cohort of 50,000 subjects, the Collaborative Perinatal Project (CPP) (12), which agreed to contribute to the I4C.

PROGRESS IN POOLING DATA WITHIN I4C

From the time of its establishment, the intention of the I4C has been to increase power through the pooling of both questionnaire data and biospecimens for analysis wherever possible. It was established that questionnaire data was available in each domain of interest for CC across the cohorts. In addition, most cohorts had collected some type of biospecimen. MoBa is one of five among the six more established I4C cohorts mentioned previously that are currently in a position to contribute to the pooled efforts of the I4C with both questionnaire data and biospecimens.

Questionnaire data

Opportunely, the questionnaires used in different cohorts had similarities and this may have been because of the linking role that ALSPAC played in these studies. It had developed its methods with some collaboration with TIHS then, later, there was communication between ALSPAC, MOBA and DNBC, and, in turn, NCS. Even so, the original questionnaires were generally in the language of the country concerned. This necessitated the availability of translated question-

naires and there was some cohort-to-cohort variation in aspects of different questions.

Our access to questionnaires early on in the process allowed us to collate information on key aspects such as timing of data collection (pregnancy, birth, early life), age at follow-up intervals, key domains for which data were collected (demographics, anthropometrics, lifestyle exposures, environmental exposures, dietary factors and general health) and who the collected data related to (mother, father or child). Approved questionnaires are electronically accessible to all I4C members through the NCI-I4C portal which was established in 2010 (https://communities.nci.nih.gov/i4c/default.aspx)

Harmonization and pooling of data

To this point, data has been received for pooling from (alphabetically): the Avon Longitudinal Study of Parents And Children (ALSPAC, UK) (13), the Collaborative Perinatal Project (CPP, USA) (14), the Danish National Birth Cohort (DNBC, Denmark) (15), the Jerusalem Perinatal Study (JPS, Israel) (16), the Norwegian Mother and Child Cohort Study (MoBa, Norway) (17) and the Tasmanian Infant Health Survey (TIHS, Australia) (18). In total, this represents data from 360,000 live births in the six participating cohorts, among whom, over 500 cancers have occurred. The dataset available for analysis at present includes all live births for ALSPAC, CPP, JPS and TIHS. Table 1 displays information on the timing of recruitment of these cohorts, the numbers of subjects in each, the number for whom data has been transferred and the number of cancers that have been notified to the I4C International Data Coordinating Centre so far. For MoBa and DNBC that number represents a 10% sample of the cohort. For the other cohorts data on all subjects was transferred.

Over 30 key variables have been harmonised across the six cohorts (listed above). A pooled dataset has been compiled for use in the current research proposals (see below). While complete harmonisation of all variables has not always been possible given the differences among cohorts, when the time came for pooling, decisions were made based on what could be pooled with minimal compromise of the original representation of the recorded data.

All CC cases and a random 10% sample of non-cases from MoBa and DNBC are included. The pooling and harmonisation work was undertaken at the I4C

International Data Coordinating Centre (IDCC) within the Murdoch Children's Research Institute in Melbourne, Australia, and commenced following establishment funding from the NCS and NICHD in 2008.

PROGRESS IN ANALYSIS OF DATA AND THE CONTRIBUTION OF MoBa TO CURRENT I4C RESEARCH EFFORTS

Currently, the I4C has a number of active working groups that are leading work in examining important putative exposures for CC. The examination of these hypothesised associations has been approved by the I4C Steering Committee which includes representatives from all participating cohorts). These hypotheses include:

- Birth weight (Ora Paltiel, JPS)
- Occupation exposure to pesticides and residential proximity to agricultural land (Ben Booth and Mary Ward, NCI)
- Paternal age (Anne-Marie Nybo-Andersen, DNBC)
- Birth order (Ora Paltiel)
- Prior fetal loss (Gabriella Tikellis, IDCC)
- Epigenetic changes, 'omics' and childhood cancer (see below) (Zdenko Herceg, Monica Cheng Munthe-Kaas, Akram Ghantous, IARC and Terry Dwyer)

In addition, an ecological analysis (lead by Gabriella Tikellis at IDCC) is being undertaken to explore the association between childhood cancer incidence and the prevalence of a range of putative exposures across each of the six cohorts. Cohort prevalence data from MoBa and information regarding childhood cancer cases are a key component of this work that will lead to the publication, possibly for the first time, of population-level prevalence of exposures and childhood cancer incidence rates. All of these research areas will require the analysis of data that have been harmonised and pooled.

In the first I4C research paper that was submitted for publication, the association between birth weight and childhood cancer was examined using pooled data from an initial 380,000 live births in which 547 childhood cancer had occurred. Because data was missing on an important covariate, gestation, for a large fraction of JPS subject, those with missing data on this variable were removed from the analysis. The final analysis took place on a subset of the data which included 377 cases of CC. MoBa contributed data on 101 of these cases. This paper will make an important contribution to the childhood cancer field as it provides the first prospective data not only on birth weight, but also on the association after examination of a range of important covariates such as gestational age and material size associated with birth weight and with risk of all childhood cancers, and in particular, leukaemia. For a number of these covariates the examination of their role in this causal setting has not been possible using prospective data because of the lack of that data

in record linkage studies.

Given its large sample size, the broad range of data it has available across relevant exposure domains, the availability of prospectively collected biospecimens, and the collaborative efforts and expert knowledge provided by cohort representatives, MoBa makes a significant contribution to each of the above mentioned areas of research within the I4C. It is anticipated that this contribution will continue and perhaps even be expanded as additional domains are explored as part of the I4C.

Availability and use of biospecimens

Of the six cohorts in a position to contribute to the current pooled efforts, four (MoBa, ALSPAC, CPP and DNBC) have cord blood serum and DNA available. TIHS has stored neonatal blood spots from which some DNA can be extracted. Because such stored blood spots are likely to be available in some future cohorts where stored DNA from whole blood will be difficult to obtain, work in the laboratory to determine how useful these cards will be as a source of DNA for genomic and epigenomic analysis has been needed. To contribute to an understanding of this I4C, in collaboration with International Agency for Research on Cancer (IARC), used DNA extracted from neonatal blood spots from TIHS and also from NCS, to determine how well each type of biospecimen would support epigenetic investigation via whole methylome analysis (19). DNA methylome analysis represents one of the omics approaches in which the quality of extracted DNA has to be optimum, specifically because methylomics often involves stringent pre-processing steps of bisulfite conversion that can degrade DNA. Blood spots tested from some I4C cohorts yielded DNA of sufficient quantity and quality using protocols optimized for this purpose and, to permit valid locus-specific and epigenomewide DNA methylation analyses (19). Locus-specific methylation was analyzed using pyrosequencing while methylome-wide profiling was done using Infinum HumanMethylation450 (HM450) arrays (Illumina Inc.), which was the method of choice also tested on cord blood DNA from the MoBa cohort (described later). HM450 enables the detection and quantitation of DNA methylation levels at 486,685 CpG sites across the genome and represents one of the most comprehensive microarray methods to date for investigating the methylome (20). The results of this work were the subject of the first I4C publication on biospecimens (19).

Of the four cohorts that have stored cord blood DNA (ALSPAC, CPP, DNBC and MoBa), MoBa was the first to provide IARC with DNA from this source for epigenetics analysis. Quantified cord blood DNA from cases and controls, matched by birth year, were extracted, identified, linked to clinical phenotype and exposure data, and shipped to IARC for epigenetic analysis involving SNPs across the genome. This 'whole methylome' analysis requires high quality

DNA because, as mentioned previously, such profiling requires pre-processing steps that include bisulfite conversion which can degrade DNA. DNA from the MoBa biospecimens was tested at IARC for methylome-wide profiling using HM450 arrays, and 99.3% of samples passed all HM450 quality controls. Analysis linking early-life exposure, epigenetics and childhood cancer risk is currently being undertaken by MoBa, IARC and I4C.

Future involvement of MoBa data and biospecimens in a complex 'omics' analysis of the relationships between the 'exposome', biological systems, and childhood cancer

The evolution of new technologies that can be used with small quantities of biospecimens on large numbers of subjects to either improve measurement of environmental exposures or to characterise biological processes that might mediate effects on disease has opened up new opportunities for examining potential exposure-disease relationships in cohort studies. These technologies include the measurement of the metabolome (21), epigenome (22), trancriptome (9), and what might be termed the infectome (19). They can together be referred to under an omics' rubric.

The exploitation of these new omics methodologies has the potential to strengthen the evidence base that supports the role of putative environmental factors, and to gain a greater understanding of how they might operate in childhood cancer. Optimally, the fabric of evidence would be constructed using questionnaire data and biospecimens collected concurrently on the same subjects as is the case in the cohorts comprising I4C.

Recognising the possibilities that this research platform presents for confirming putative causal pathways of environmental origin, or for identifying novel pathways, our intention is to use the prospective questionnaire data collected either during pregnancy or at birth, along with cord blood biospecimens, from those cohorts that can currently contribute to focus on the role of fetal exposures. The cohorts that can contribute both questionnaire data and cord blood specimens and have indicated a desire to participate in this exciting new effort include MoBa, DNBC and the Japan Environmental and Children's Study (JECS) (23). From the approximately 300,000 subjects in these three cohorts

we expect approximately 700 cases of CC. Because of cost, and a wish to utilise the smallest amount of the biospecimen resource possible, cohort controls on whom biospecimens will be sought sampled on a 2:1 ratio to cases. Using both DNA and serum from cord blood we will undertake extensive analyses of the epigenome, the adductome, the metabolome, the transcriptome, and the infectome. We hope to undertake genomic analysis, focused on chromosomal translocations in some cohorts. The investigation will include the testing of a priori hypotheses as well as broad based hypothesis free analyses. We propose that it is not unreasonable to anticipate a 'leap forward' in knowledge concerning the aetiology of childhood cancer through the accomplishment of this investigation.

Table 2 provides some basic information on the cohorts that will be involved in this strategy and a list of new cohorts that we hope will contribute additional data in the years to come.

It is evident that MoBa is central to I4C's current strategies to analyse both pooled questionnaire data and biospecimens to gain new insights into preventable causes of CC. It is providing approximately 25% of the information we are now analysing from the questionnaire database and has indicated its intention to be one of three large birth cohorts that will provide the data for the exciting new 'omics' study. We are hopeful that the yield from the work to which MoBa has contributed so much will justify the contribution so many people have made to the effort.

Table 2. I4C cohorts – available cord blood biospecimens on cancer cases.

Existing cohorts	No. subjects	Cord bloods
DNBC	100,000	250
JECS	100,000	200
MoBa	100,000	250
SUBTOTAL	300,000	700
Cohorts at an earlier stage	No. subjects	Cord bloods
France	16,000	40
UK	100,000	250
NCS-USA	100,000	250
Australia	100,000	250
Brazil	100,000	250
TOTAL	716,000	1,740

References

1. Steliarova-Foucher E, Stiller C, Kaatsch P, et al. Geographical patterns and time trends of cancer incidence and survival among children and adolescents in Europe since the 1970s (the ACCISproject): an epidemiological study. *Lancet* 2004;**364**(9451):2097-105.
2. Magrath I, Steliarova-Foucher E, Epelman S, et al. Paediatric cancer in low-income and middle-income countries. *Lancet Oncol* 2013;**14**(3):e104-16.
3. Ries L, et al. *Cancer incidence and survival among children and adolescents*: United States SEER program 1975-1999, 1999.
4. Greaves MF, Wiemels J. Origins of chromosome translocations in childhood leukaemia. *Nat Rev Cancer* 2003;**3**(9):639-49.

5. Spector LG, Ross JA. Infant leukemia: finding the needle in the haystack. *Cancer Epidemiol Biomarkers Prev* 2006;**15**(12):2331.
6. Brown RC, Dwyer T, Kasten C, et al. Cohort profile: the International Childhood Cancer Cohort Consortium (I4C). *Int J Epidemiol* 2007;**36**(4):724-30.
7. Caughey RW, Michels KB. Birth weight and childhood leukemia: a meta-analysis and review of the current evidence. *Int J Cancer* 2009;**124**(11):2658-70.
8. Wild CP. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev* 2005;**14**(8):1847-50.
9. Rappaport SM, Smith MT. Environment and disease risks. *Science* 2010;**330**(6003):460-1.
10. Wild CP, Scalbert A, Herceg Z. Measuring the exposome: a powerful basis for evaluating environmental exposures and cancer risk. *Environ Mol Mutagen* 2013;**54**(7):480-99.
11. Rappaport SM. Biomarkers intersect with the exposome. *Biomarkers* 2012;**17**(6):483-9.
12. Rappaport SM, Li H, Grigoryan H, Funk WE, Williams ER. Adductomics: characterizing exposures to reactive electrophiles. *Toxicol Lett* 2012;**213**(1):83-90.
13. Boyd A, Golding J, Macleod J, et al. Cohort Profile: the 'children of the 90s' – the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol* 2013;**42**(1):111-27.
14. Klebanoff MA. The Collaborative Perinatal Project: a 50-year retrospective. *Paediatr Perinat Epidemiol* 2009;**23**(1):2-8.
15. Olsen J, Melbye M, Olsen SF, et al. The Danish National Birth Cohort – its background, structure and aim. *Scand J Public Health* 2001;**29**(4):300-7.
16. Harlap S, Davies AM, Deutsch L, et al. The Jerusalem Perinatal Study cohort, 1964-2005: methods and a review of the main results. *Paediatr Perinat Epidemiol* 2007;**21**(3):256-73.
17. Magnus P, Irgens LM, Haug K, Nystad W, Skjærven R, Stoltenberg C. Cohort profile: the Norwegian Mother and Child Cohort Study (MoBa). *Int J Epidemiol* 2006;**35**(5):1146-50.
18. Dwyer T, Ponsonby AL, Blizzard L, Newman NM, Cochrane JA. The contribution of changes in the prevalence of prone sleeping position to the decline in sudden infant death syndrome in Tasmania. *JAMA* 1995;**273**(10):783-9.
19. Ghantous A, Saffery R, Cros MP, et al. Optimized DNA extraction from neonatal dried blood spots: application in methylome profiling. *BMC Biotechnol* 2014;**14**:60.
20. Bibikova M, Barnes B, Tsan C, et al. High density DNA methylation array with single CpG site resolution. *Genomics* 2011;**98**(4):288-95.
21. Psychogios N, Hau DD, Peng J, et al. The human serum metabolome. *PLoS One* 2011;**6**(2):e16957.
22. Herceg Z, Vaissiere T. Epigenetic mechanisms and cancer: an interface between the environment and the genome. *Epigenetics* 2011;**6**(7):804-19.
23. Kawamoto T, Nitta H, Murata K, et al. Rationale and study design of the Japan environment and children's study (JECS). *BMC Public Health* 2014;**14**:25.