EUCCONET Record Linkage Workshop

EUCCONET Data Management Interest Group

EUCCONET International Workshop 15th – 17th June 2011

Organisers:

- Andy Boyd, ALSPAC Data Linkage Manager, ALSPAC
  (University of Bristol, UK)
- Patricia Schreuder, MoBa Cohort Manager, MoBa
  (Institute of Public Health, Norway)

Venue: Norwegian Institute of Public Health, Kalfarveien 31, Bergen, Norway.

Scientific Report

Summary

The administration and investigative use of child cohort studies requires an enormous investment of time and resources, from funders, the research community and study participants. The scientific potential for a cohort study is maximised through the collection of broad, accurate and complete data. In most studies the scientific investigator needs to balance the advantages of data collection against limited financial resources. Administrative records have been used to maximise the ability of cohort studies to identify the causal influences on physical and psychosocial health, and thus help to ensure the scientific return on existing investment and cost efficient means of future data collection. The use of these records has increased as new technologies allow access to a broader range of electronically held records.

This workshop brought together researchers and study administrators from EU child cohort studies along with specialist speakers whose work is closely connected with linkage and cohort studies. The rationale for the workshop was to enable linkage practitioners and experts to share their experiences and provide insight into the attributes that can lead to successful linkage projects.

The workshop was designed with three distinct approaches to share and learn from each other's linkage strategies and methodologies:

1. To encourage the sharing of best practice and the development of a linkage focused network between primarily EU cohorts. There were representatives from 11 longitudinal child

European Science Foundation — SETTING SCIENCE AGENDAS FOR EUROPE

EUROPEAN CHILD EUCCONET COHORT NETWORK

University of BRISTOL

Norwegian Institute of Public Health

cohorts based in Europe with representatives from non-EU countries (1 from New Zealand and 1 from the USA) providing a wider perspective.
2. To provide summary training to enhance the understanding amongst the delegates of key methodologies in the field, including references to introductory texts and key publications.
3. To provide a forum to illustrate the work of some exemplar projects that demonstrate the common issues faced by linkage researchers within EU cohort studies.

The principal impact of this workshop was to discuss and share the best practice found across a range of EU cohort studies. Participants were seen to contribute to and take away new information which can be used to evaluate and shape data linkage strategies within each individual study.

Scientific Background

Comprehensive and in-depth administrative records are routinely collected and stored electronically on EU citizens. Social data include records on income, employment and other aspects of social position, data on children's performance in educational assessments, data on the physical environment and involvement in criminal and antisocial activity. Health data include the records from both primary and secondary care as well as national population registries (including birth, death and cancer registries).

The challenge for cohort studies is to develop robust mechanisms that are secure, legal and ethically suitable for cohort members. There is a wide variation in the scope and access requirements needed to obtain data across Europe from the integrated social registries of Scandinavia to those developed by government departments with bespoke databases where access is heavily dependent on bureaucratic and under-developed procedures. However legislative requirements, including the EU Data Protection legislation, are applicable to all EUCCONET studies. This workshop provides a forum for linkage specialists and researchers working in, or commissioning this work to meet and discuss the potential and best practice found within the EUCCONET studies.

1. Study Reports

Representatives from ten countries presented brief introductions on their studies and described the linkage based data collection strategies they had in place or planned for the future. The scope of their linkage activities has been summarised in table 1.

a. MoBa, Norway.
Nina Hagesæther summarised MoBa as a health focused cohort of 100,000 pregnancies recruited from across Norway between 1999 and 2009. Linkage is aided in Norway by the availability of a unique 'citizen' personal ID number issued at birth which is common across all registries and datasets. This ID number was collected along with participant consent during study enrolment, allowing efficient and accurate linkage to contact details and phenotypic data. While there are few technical barriers to accessing linkage data there are strict data access requirements.

b.   Generation R, Netherlands.

Claudia Kruithof summarised Generation R as an urban, multi-ethnic birth cohort of 10,000 pregnancies recruited in Rotterdam, Netherlands between 2001 and 2006. Generation R is a health focused study. Within the Netherlands there are few record linkage possibilities due to limited national registries, with restricted access. Data are either stored anonymous or with different identifiers within the healthcare system. Primary and secondary care data are stored locally with use of different identifiers. Access requires written consent from the study participant. The study has had success in linking to local primary care information but data quality and limited practices using the software meant a low level of successful linkage. Data collectors have found benefits from collecting data using standardised forms in health centres, including health ID numbers.

c.   ELFE, France.

Xavier Thiery summarised ELFE as a multi-discipline, nationally representative cohort of 20,000 children born in 2011. The study aims to test many environmental exposure related hypothesis so linkage to date concentrates on geo-matching to environmental data, census data and healthcare usage through the French social security system. ELFE found a benefit in collecting personal ID numbers during recruitment; these link both the parent and child records. French routine records are only kept for 3 years and access requires participant consent. The limited history of data collection in France using linkage is causing barriers to data access.

d.   Growing Up in New Zealand, New Zealand.

Peter Tricker summarised Growing up in New Zealand as a representative, multi-ethnic cohort of ~7,000 children recruited in Auckland during 2009-10. This multi-discipline study is using linkage to collect perinatal data from state and private organisations. They have a high consent rate and a high match rate to the health individual ID number. Considerable effort has been required though to clean and process linkage data due to inconsistent approach from healthcare providers. This process is easier from state health providers with consistent data collection and storage practices.

e.   ALSPAC, England.

Kerry Humphries summarised ALSPAC as a regionally based cohort of 14,500 recruited in pregnancy during 1990-1992. ALSPAC is health focused but recent move towards multi-disciplinary questions. Linkage has been established to education data and mortality and cancer registries. Currently consenting study children for linkage to health, education, benefits and earnings and criminal records. Encountering different problems and technical challenges for each data set as all are based in different government departments with no common linkage IDs or data access procedures and each has a different ethics framework. The linkage pilot study used randomised controlled trial methods to test different approaches, this found that a professionally designed pack increased the speed of participant response and that telephone based reminders were more effective than postal reminders.

f.   GINI/LISA, Germany.

Wolfgang Lohr summarised both projects as regionaly recruited health studies. GINI has a cohort of ~6,000 recruited in 1995-1998 and LISA has a cohort of ~3,000 recruited in 1997-1999. Wolfgang raised the important issue of linkage within studies, discussing the need to standardise and link data across multiple study centres and different data collection methods. These processes include linkage techniques, such as using multiple personal identifiers to assess the plausibility of each data join rather than relying on a study ID number alone. The studies have developed innovative methodologies to use Google products to collect geo-coded data and validate address data but they have substantial concerns regarding data ownership and the potential for Google to retain sensitive study data.

g. Born In Bradford, England.
    David Bolus summarised Born in Bradford as a health focused cohort recruited from the city of Bradford, England. The sample can be characterised as urban and multi-ethnic. The study is embedded in the National Health Service (NHS) which provides advantages to accessing health data at every level; maternity, primary care and secondary care. The study has an additional advantage in that almost all local general practices use the same software package. This aids data collection and data processing. The study considers strong engagement activities, with a local community focus, with the health community and study participants as a key part of their success.

h. Millennium Cohort, U.K.
    Heather Joshi described the Millennium cohort as a multidisciplinary cohort of ~19,000 recruited across the U.K during the year 2000. Permission to link to records is consent based and currently extends to cancer and mortality registries, education records and geo-coded datasets. Collection of data from hospital episode statistics and primary care records is in development. The consent was conducted during interviews at age 14 with good results (Child: 95%, Main parent 90% , Partner 86%). Heather raised discussion points regarding confidentiality in linkage, data ownership and the use of 'third parties' to establish the links, additionally the study are having to identify ways of managing consent over time in terms of who gives consent and how long it lasts for.

i. Environments for Healthy Living, Wales.
    Sinead Brophy summarised the EHL as a traditional birth cohort study embedded within a total population electronic cohort study (the Wales Electronic Childrens Cohort or WECC). While EHL is a new study with a small enrolled pilot sample of 300 families the WECC contains data on ~800,000 children from 1990 onwards. The linkage between the two will combine phenotypic data from EHL with administrative records from WECC into a single anonymised data set. The linkage is conducted using NHS patient ID number and the study sought participant consent. Sinead described the advantages this system offers in terms of data breadth, ease of access and also a reduced participation burden on the cohort.

j. NLSY79, USA (Paula gave a verbal presentation so there are no slides available).
Paula Baker summarised the NLSY79 as a multi-centre urban study of around 12,000 participants recruited at age 14-22 in 1979. The sample is nationally representative and uses linkage to collect contextual data about employment and through geo-coding to census information. The ADD Health study has identified friendship networks in a school context.

Table 1: Current and Planned Linkage Activities by individual Cohort Study.

Linkage Data Category		ALSPAC	Born In Bradford	ELFE	Environments for healthy living	Generation R	GINI LISA	Growing up in New Zealand	Millennium Cohort	MoBa	NLSY79
Health	Birth Registry	A	A	N	A	A	N	A	D	A	N
	Maternity Records	A	A	N	A	A	N	A	D	A	N
	Cancer Registry	A	D	N	P	N	N	N	A	A	N
	Death Registry	A	D	N	P	N	N	A	A	A	N
	Hospital Episodes (secondary care)	D	D	N	P	N	N	A	A	A	N
	Doctors Records (primary care)	D	A	N	P	N	N	N	D	N	N
	Prescriptions	D	A	D	P	P	N		D	A	N
Education	Exam Results	A	P	N	P	N	N	N	A	N	N
	School level data	A	P	N	P	D	N	P	A	N	N
	University (higher education) data	A	P	N	P	N	N	N	A	N	N
Employment	Employment Records	D	N	N	N	N	N	N	D	N	N
	Salary (Earnings)	D	N	N	N	N	N	N	D	N	N
Social Services	Government financial support (housing and income benefits)	D	N	D	N	N	N	N	D	N	N
	Social care	N	N	N	P	N	N		N	N	N
Criminal Activity	Criminal Convictions	D	N	N	N	N	N	N	N	N	N
Geographical Linkages	Location data	A	N	N	A	N	A	D	A	N	N
	Environmental Exposures	A	N	A	N	N	N	D	N	N	N
Other											
	Participant Address Data	A				A				A	
	Friendship Networks	A									A
Key: A 'Active', D 'In development', P 'Planned for the future', N 'Not planned'											

Group discussion following these presentations centred on the differing approaches to data access, consent requirements and different cultural approaches to data access from different data owners. The Norwegian example of a centralised registry based system with a single identifier was seen as the most desirable system with all other studies facing difficulties in terms of consent, technical barriers to data or facing barriers such as risk aversion, lack of data sharing processes or data

cleanliness issues when dealing with data owners. However it is important to note that while having few technical issues MoBa still required consent and ethical clearance to access routine records.

These access issues are compounded after data confidentiality breaches, potentially in terms of willingness to consent but also in terms of willingness of data owners to provide access to data. The resource available to the EHL study in Wales provides the closest comparison to Norway but relies on a complex technological solution provided by the SAIL infrastructure. This model was described in the next session.

Record linkage methodologies are applied using a range of techniques, the choice of which is frequently driven by necessity and controlled by factors such as data access, data quality and the availability of universal citizen Identification Numbers (IDs). These data issues frequently necessitate probabilistic methodologies.

2. Expert Speakers

a. An introduction to probabilistic Linkage Methods

The opening paper authored by 'Mac' McDonald was taken from a training programme offered by his research unit and provided participants with both a summary understanding of linkage methodologies and references to key methodology and exemplar references. The paper defined deterministic and probabilistic methodologies and described how they differed and the circumstances in which each were appropriate. The paper introduced the Felligi and Sunter (1969) probabilistic model and its principal concepts[1]. The paper highlighted key texts[2] and references which described quality issues[3-6, 8-10] and relevant exemplars of the linkage methodologies. Mac also discussed an open-source software package[7] that can be used to facilitate probabilistic linkage.

b. Privacy Protection and the SAIL Databank

David Ford presented a summary of the Secure Anonymised Infrastructure for Linkage (SAIL) databank[12]. This resource provides researchers with over 1 billion individual records collated from multiple routine health and administrative records from the Welsh national population. The privacy protection techniques and the use of an independent third party to link the records has allowed the research team to convince the data owners to approve access to all records without patient consent. This is legal under E.U. and U.K. data protection legislation as anonymity is ensured. David provided examples of how nested samples from SAIL, such as WECC can interact with 'traditional' cohort models, the EHL study is an example of this. To establish this resource the team faced substantial technical barriers in terms of data quality and access 'pipelines' and also in developing robust governance arrangements, including a data access committee, that satisfied the data owners.

c. Personal Identity Protection Solutions

Kerina Jones outlined a methodology to protect participant confidentiality by moving from de-identifying a data set to 'anonymising' a data set through ensuring the removal of small cell counts and an individual's records having a unique combination of values. There are various techniques that data managers employ to control for the risk of an individual's identity being disclosed, and

therefore breaching their confidentiality. These techniques, including suppression, aggregation and perturbation reduce the ability of the dataset to inform research analysis. Establishing the balance between data utility (the power of the data to inform research) with participant confidentiality is commonplace within linkage community as cohort studies will need to make commitments regarding this to the data owners. Jones outlined a series of algorithms that can be used to achieve k-anonymisation[14 15] within a dataset. K-anonymisation ensures a given record cannot be distinguished from at least (*k* - 1) other records. The various algorithms and tools described can be used to help researchs strike an optimal balance between utility and confidentiality.

d. The NHS Information Centre (NHS IC) & Information Governance Controls

Clare Sanderson described the work of the NHS IC in centralising and facilitating access to NHS medical data in the United Kingdom. This work was presented as an example of best practice as how a government department can take the role of 'honest broker' and provide a centralised team to enable linkage and facilitate access to their routine data. Functions the NHS IC take on include undertaking linkage between datasets, cleaning participants personal identifiers (names and dates of birth), advising on consent specifications and guiding researchers through ethical review. The NHS ICs remit also includes establishing and enforcing information governance controls within the NHS. They have developed a series of policies and training known as the Information Governance Toolkit[13]. One of these protocols describes a best practice methodology, known as 'Data Safe Havens', of separating personal identifiers from research data and ensuring that these data (in this case linkage data) are processed and de-identified as soon as possible. It has been recommended in the UK that cohort studies adopt these principles[16].

3. Exemplar Research

a. Patterns for consent in the Millennium Cohort Study

Sosthene Ketende presented a paper describing how consent rates for linkage in the Millennium Cohort varied across the four countries in the United Kingdom (Wales, Scotland, Northern Ireland and England). When seeking consent for health, education and economic linkage the researchers found regional/national variation in consent rates and that ethnicity, age, education and lone parent status were predictive of consent[17 18].

b. Young People's perspectives on record linkage

Lindsey Brown presented findings from qualitative interviews on ~50 young people eligible to take part in ALSPAC. The interviews were in depth and sought the opinions of individuals from a range of socio-economic backgrounds and those with differing histories of participation in ALSPAC. A common thread between the majority of interviewees was that the concept of record linkage was new to them and a detailed discussion of the topic took time during which the participants views could change dramatically when presented with different scenarios. Lindsey presented evidence illustrating how the views of individuals regarding consent and linkage activities varied from not wanting linkage to occur, through the necessity of consent or being given full information to some individuals were happy for linkage to occur regardless of consent or notification. These positions

were influenced by the individuals trust and perception of the study, their data history and concerns over how the data of others was to be used.

c. The family based longitudinal database of Norwegian citizens

Dominic Hoff presented a paper on his work in retrospectively defining a multi-generational cohort from census records. This methodology allowed the creation of a dataset with sufficient size to allow the formal statistical testing, such as difference between paternal and maternal effects, grandparental effects, within and between sibship-comparisons, and associations between genetic variants and later disease. This sample could be linked to the death registry, census variables, birth registry and in some cases the Cohort of Norway (CONOR) datasets. This approach has led to the creation of a resource with a large sample size with onward linkage to phenotypic datasets. The process was complicated by difficulties in defining the biological father figure from historical records.

d. MoBa, Intergenerational Linkage

Arild Sunde described the MoBa administrative database system and how it interacts with the Norwegian registry to collect data. Each Norwegian citizen is issued a unique ID number which includes their date of birth, this information can be used to check validity. Arild described how MoBa used deterministic linkage to expand the scope of their data collection to grandparents and siblings. Data were collected from a range of registries. The principle difficulties lay with gaining ethical approval. Moba informed the cohort of their intentions and received 35 objections from the cohort. All data are pseudonymised prior to research use.

Rolv Skjærven discussed how this resource has been used to look at intergenerational comparisons between outcomes including birthweight, reproduction and if these are true intergenerational affects or simply due to social confounding [19][20].

e. Identifying an adolescent peer network within ALSPAC using probabilistic linkage

Andy Boyd presented a paper describing how probabilistic linkage had enabled the identification of a friendship network within the ALSPAC cohort study[11]. While there have been previous studies which have identified friendship networks these have all used closed communities such as schools or prisons. ALSPAC instead opted to send questionnaires to participants and asked them to identify their 5 closest friends. Using probabilistic methods proposed by Felligi and Sunter[1] ALSPAC were able to de-duplicate the responses (n=3,000 respondents), identifying the common entries amongst the nominated friends(~11,000 individuals from ~14,000 nominations). In a second linkage exercise they were able to match these nominated friends against the ALSPAC administrative database to establish which were in the database (~4,500 individuals).

Scientific Impact and Future Directions

The workshop aims were: 1.to facilitate the networking and sharing of study level practice amongst EU child cohort studies, 2. to provide summary level training of record linkage methodologies, backed with exemplar research applications, and 3. To discuss issues of consent, participant

understanding and privacy protection through the presentation of exemplar research and from expert speakers.

1. Networking and sharing of best practice

The study update session facilitated introductions and networking, allowing participants to discuss and share the best practice regarding linkage activities found across a range of EU cohort studies. Participants were seen to contribute to the discussion and take away new information which can be used to evaluate and shape data linkage strategies within each individual study. Due to the breadth of study designs represented we were able to consider linkage applications from a wide range of perspectives.

2. Introduction to probabilistic linkage theory

The workshop participants were drawn from studies of all types and stages of implementation. While some study representatives had considerable working understanding of linkage methodologies others were at a planning stage or had access to national registries which systematically include individual level 'linkage'. To address this range of experiences the organisers incorporated papers designed to provide introductory level training on the history[1] and principles of linkage methodologies[2] alongside exemplar papers[11][12] of linkage in practice. The McDonald paper provided participants with both a summary understanding of linkage methodologies and references to key methodology and exemplar references. This understanding was reinforced by exemplar papers of cohort based research based on these methodologies. The organisers received feedback that this summary was considered valuable and gave participants the understanding and resources to see the extended possibilities allowed by probabilistic linkage.

> **Feedback Question:** What was the most useful presentation(s)/session(s) for you, and why was it so useful?
>
> **Feedback Answer:** *Session on probabilistic matching. Currently we use deterministic matching to link to health data but in a couple of years we will most likely want to match to education records and will have to use probabilistic matching if the national student identification system is not in place for pre-schoolers*

The training objective was achieved, with the first and last paper of the workshop focused on the key Fellegi and Sunter probabilistic linkage model.

3. Participant Understanding and Consent & Privacy Protection

The papers presented by Ford, Brown, Jones, Ketende and Sanderson outlined varying aspects of privacy protection, information governance and the perspectives from the point of view of the research participant. The group discussed the importance of the considerations and how cohort understanding and perceptions interact with may influence differing consent rates between linkage data sources and across different populations. Brophy reported that a childrens focus group from the EHL study considered the study to form part of their NHS treatment and as such didn't require consent, however they did expect to be informed about the study and this use of their information.

4. *Future Directions*

There was agreement within the group regarding the importance of data collection via record linkage within a cohort study. It is hoped that enthusiasm from the workshop will lead to onward networking and dissemination of best practice with the aim of facilitating cross study research utilising linkage data. It was noted that there is considerable variation across EU countries in terms of the availability and conditions of access, the paper from Elias encouraged the delegates to join in the debate and described possibilities of how cohort studies could influence EU debate in this area.

References

1. Fellegi IP & Sunter AB (1969) A theory for record linkage. Journal of the American Statistical Association 64, 1183-1210

2. Herzog TN, Scheuren FJ & Winkler WE (2007). Data quality and record linkage techniques. New York: Springer. 234 pp.

3. Gomatam S et al (2002) An empirical comparison of record linkage procedures, Statistics in Medicine, 21, 1485-1496 http://nisla05.niss.org/dgii/presentations/gomatam-sic-200305.pdf

4. Roos LL, Walld R, Wajda A, Bond R, Hartford K (1996) Record linkage strategies, outpatient procedures, and administrative data. Medical Care 34, 570–582

5. Jamieson E, Roberts J, Browne G (1995) The feasibility and accuracy of anonymized record linkage to estimate shared clientele among three health and social service agencies. Methods Inf Med. 34, 371–377

6. Cibella N et al Sharing Software for Record Linkage: The RELAIS Software and the Italian and Spanish Experience. Found at (last accessed 30 September 2011): http://epp.eurostat.ec.europa.eu/portal/page/portal/research_methodology/documents/S7P2_SHARING_SOLUTIONS_FOR_RECORD_LINKAGE_CIBELLA_TUOTO_.pdf

7. Record Linkage at ISTAT (RELAIS) Software: Found at (last accessed 30 September 2011): http://www.istat.it/it/strumenti/metodi-e-software/software/relais

8. Howe GR (1998) Use of computerized record linkage in cohort studies. Epidemiologic Reviews 20, 112–121

9. Krewski D, Dewanji A, Wang Y, Bartlett S, Zielinski JM & Mallick R (2005) The effect of record linkage errors on risk estimates in cohort mortality studies. Survey Methodology 31,13–21

10. Brenner H, Schmidtmann I & Stegmaier C (1997) Effect of record linkage errors on registry-based follow-up studies. Statistics in Medicine 16, 2633–2643

11. Burgess et al (2001). School ties: An analysis of homophily in an adolescent friendship network. CMPO Working Paper 11/267. Found at (Last accessed 30 September 2011): http://www.bristol.ac.uk/cmpo/publications/papers/2011/wp267.pdf

12. Lyons et al (2009). The SAIL databank: linking multiple health and social care datasets. *BMC Medical Informatics and Decision Making* 2009, 9:3 1472-6947

13. National Health Service (NHS) Information Governance Toolkit. Found at (last accessed 30 September 2011): https://www.igt.connectingforhealth.nhs.uk/

14. El Emam et al (2009). A globally Optimal k-anonymity Method for the De-Identification of Health Data. *J Am Med Inform Assoc 2011;18:3-10*

15. Pierangela Samarati and L. Sweeney. *k*-anonymity: a model for protecting privacy. *Proceedings of the IEEE Symposium on Research in Security and Privacy* (S&P). May 1998, Oakland, CA.

16. Walport M, Thomas R (2008) Data Sharing Review Report. Found at (last accessed 30 September 2011): http://www.justice.gov.uk/reviews/datasharing-intro.htm

17. Hockley C et al (2007) Linking Millennium Cohort data to birth registration and hospital episode records. Paediatric and Perinatal Epidemiology 22, 99-109

18. Tate R et al (2006) Mother's consent to linkage of survey data with her child's birth records in a multi-ethnic national cohort study. International Journal of Epidemiology 35, 294-298

19. Lie RT, Wilcox AJ, Skjaerven R. Maternal and paternal influences on gestational age. *Obstet Gynecol* 107:880-85, 2006

20. Skjaerven R (2005). Recurrance of pre-eclampsia across generations: exploring fetal and maternal genetic components in a population based cohort. BMJ 331:877

EUROPEAN SCIENCE FOUNDATION — SETTING SCIENCE AGENDAS FOR EUROPE

EUROPEAN CHILD EUCCONET COHORT NETWORK

University of BRISTOL

Norwegian Institute of Public Health

EUCCONET Record Linkage Workshop

EUCCONET Data Management Interest Group

Bergen 15th – 17th June 2011

Venue: Norwegian Institute of Public Health, Kalfarveien 31, Bergen, Norway.

15th June: Evening Introductory Session (18:00 – 20:00)

(light snack on arrival)

18.00 – 18.20   Workshop Welcome: Aims for the workshop

Andy Boyd (School of Social and Community Medicine, Bristol)

18.20 – 19.00   An introduction to Probabilistic Record Linkage

Presentation Author: *John 'Mac' McDonald (Institute of Education, London)*

Due to illness Mac was unable to attend and Andy Boyd presented this paper in his absence

19.00 – 20.30   Armauer Hansen and the Bergen Lepramuseet

Sigurd Sandmo

The institute hosting the workshop houses the laboratories used by Armauer Hansen in his pioneering work into the study of Leprosy. Sigurd will give a talk on Hansen, a tour of the *Lepramuseet* and illustrate the role the world's first national patient register played in understanding Leprosy.

Buffet Meal

16th June: Study Updates, Information Governance & Disclosure Control

Morning session, Chair: tbc

Coffee & Registration (from 8.40)

09.00 – 09.10   Welcome and Introduction

Reports from EUCCONET cohort studies on the role of record linkage in their project.

09.10 – 09.30   The Norwegian Mother and Child Cohort Study

Nina Hagesæther (Norwegian Institute for Public Health, Bergen)

09.30 – 09.50   Generation R

Claudia Kruithof (Erasmus Medical Center, Rotterdam)

09.50 – 10.10   ELFE

Xavier Thierry (Institut national d'études démographiques, Paris)

10.10 – 10.30   Growing Up in New Zealand

      Peter Tricker (University of Auckland, Auckland)

10.45– 11.05   ALSPAC

      Kerry Humphries (School of Social and Community Medicine, Bristol)

11.05 – 11.35   GINI/LISA

      Wolfgang Lohr (Helmholtz Zentrum, Munich)

11.35 – 11.55   Born in Bradford

      David Bolus (NHS Bradford & Airedale, Bradford)

12.10 – 12.30   Millennium Cohort

      Heather Joshi (Institute of Education, London)

12.30 – 12.50   Environments for Healthy Living
      Sinead Brophy (College of Medicine, Swansea)

12.50 – 13.10   NLSY79

      Paula Baker (Center for Human Resource Research, Ohio)

Lunch

Afternoon session, Chair: tbc

Information Governance & Disclosure Control: One of the principal challenges of establishing record linkage data collection is to satisfy the legal and ethical concerns surrounding information governance and disclosure control.

 14.00 – 14.40   Privacy Protection & the SAIL Databank
      David Ford (College of Medicine, Swansea)
 14.40 – 15.20   Personal identity protection solutions in the presence of low copy number fields
      Kerina Jones (College of Medicine, Swansea)

Break

15.40 – 16.20   Information Governance Controls - Balancing Public Assurance and Research
      Clare Sanderson (NHS Information Centre, Leeds)
 16.20 – 17.00   Record linkage: the need for more effective processes
      Peter Elias (University of Warwick)

17.00 – 17.30   Information Governance & Disclosure Control Group Discussion

Workshop Evening Meal

17th June: Consenting & Applied examples of Record Linkage

Morning Session, Chair: Dominic Hoff

Coffee (from 9.00)

09.20 – 09.30   Welcome and Introduction

Consent for Linkage: Speakers will present evidence on study participants understanding of linkage, views about consent and consent patterns from consent campaigns. Followed by an open discussion on seeking consent for linkage.

09.30 – 10.10   Title to be provided

Sosthene Ketende (Institute of Education, London)

10.10 – 10.50   Title to be provided

Lindsey Brown (School of Social and Community Medicine, Bristol)

10.50 – 11.05   Linkage Consent & Participant Perceptions: Group Discussion

Break

Closing Session, Chair: tbc

Applied Linkage in Cohort Studies: Speakers will present examples of linkage applications in cohort studies.

11.20 – 12.00   The family based longitudinal database of Norwegian citizens

Dominic Hoff (Norwegian Institute for Public Health, Oslo)

Lunch

13.00 – 13.40   Title to be provided

Arild Sunde (Norwegian Institute for Public Health, Bergen)

13.40 – 14.20   Identifying an adolescent peer network within ALSPAC using probabilistic linkage

Andy Boyd (School of Social and Community Medicine, Bristol)

14.20 – 15.00   Closing Group Discussion

Workshop Close at 15:00

Delegate Contact Details

Alet Wijga National Institute for Public Health and the Environment P.O.Box 1 3720 BA Bilthoven The Netherlands alet.wijga@rivm.nl	**James Williams** ESRI Sir John Whitaker Sq, Dublin Ireland james.williams@esri.ie	**Peter Elias** University of Warwick Gibbet Hill Rd Coventry West Midlands CV4 7AJ UK Peter.Elias@warwick.ac.uk
Andrew Boyd ALSPAC University of Bristol Oakfield House Oakfield Grove Clifton, Bristol BS8 2BN UK a.w.boyd@bristol.ac.uk	**John Macleod** ALSPAC University of Bristol Canynge Hall 39 Whatley Road Bristol BS8 2PS UK John.MacLeod@bristol.ac.uk	**Peter Tricker** University of Auckland PO Box 18288 Auckland 1743 NZ p.tricker@auckland.ac.nz
Arild Sunde Norwegian Institute of Public Health Department for Project Support Kalfarveien 31 N-5018 Bergen Norway Arild.Sunde@fhi.no	**Julian Walls** Bradford Institute for health research Duckworth Lane Bradford BD9 6RJ UK Julian.Walls@bradfordteachinghospitals.nhs.uk	**Rachel Rosenberg** Institute of Education 20 BEDFORD WAY London WC1HAL UK r.rosenberg@ioe.ac.uk
Charlotte Birke Norwegian Institute of Public Health Division for Epidemiology N-0403 Oslo Norway charlotte.birke@fhi.no	**Kate Angel** ALSPAC University of Bristol Oakfield House Clifton, Bristol BS8 2BN UK Kate.angel@bristol.ac.uk	**Ronald van den Nieuwenhof** Erasmus MC Dr. Molewaterplein 50-60 3015 GE Rotterdam The Netherlands r.vandennieuwenhof@erasmusmc.nl
Clare Sanderson NHS 1 Trevelyan Square, Boar Lane, Leeds LS1 6AE UK clare.sanderson@ic.nhs.uk	**Kerina Jones** Swansea University Health Information Research Unit (HIRU) College of Medicine Grove Building, Singleton Park Swansea SA2 8PP Wales k.h.jones@swansea.ac.uk	**Sinead Brophy** Swansea University College of Medicine Grove Building, Singleton Park Swansea SA2 8PP Wales s.brophy@swansea.ac.uk
Claudia Kruithof Erasmus MC Dr. Molewaterplein 50-	**James Williams** ESRI Sir John Whitaker Sq,	**Sosthene Ketende** Institute of Education 20 BEDFORD WAY

60, 3015 GE Rotterdam The Netherlands c.kruithof@erasmusmc.nl	Dublin Ireland james.williams@esri.ie	London WC1HAL UK s.ketende@ioe.ac.uk
Daniel Bela NEPS/University of Bamberg Wilhelmpatz 3 Bamber Germany daniel.bela@uni-bamberg.de	**Kerry Humphries** ALSPAC University of Bristol Oakfield House Clifton, Bristol BS8 2BN K.humphries@bristol.ac.uk	**Susan Nunn** National Centre for Social Research (NatCen) 35 Northampton Square London EC1V 0AX UK susan.nunn@natcen.ac.uk
David Bolus NHS Bradford and Airedale Douglas Mill, Bowling Old Lane Bradford BD5 7RJ UK david.bolus@bradford.nhs.uk	**Lindsey Brown** ALSPAC University of Bristol Oakfield House Clifton, Bristol BS8 2BN Lindsey.Brown@bristol.ac.uk	**Wolfgang Lohr** Helmholtz Zentrum München, GmbH Ingolstädter Landstraße 1 D685764 Neuherberg Germany wolfgang.lohr@helmholtz-muenchen.de
David Ford Swansea University Health Information Research Unit (HIRU) College of Medicine Grove Building, Singleton Park Swansea SA2 8PP Wales d.v.ford@swansea.ac.uk	**Nina Hagesæther** Norwegian Institute of Public Health Department for Project Support Kalfarveien 31 N-5018 Bergen Norway Nina.hagesether@fhi.no	**Xavier Thierry** INED 133 BD dAVOUT 75980 Paris Cedex 20, France Thierry@ined.fr
Dominic Anthony Hoff Norwegian Institute of Public Health, Division for Epidemiology N-0403 Oslo Norway DominicAnthony.Hoff@fhi.no	**Ole-Martin Kvinge** Norwegian Institute of Public Health Department for Project Support Kalfarveien 31 N-5018 Bergen Norway Ole-martin.kvinge@fhi.no	
Heather Joshi Institute of Education 20 BEDFORD WAY London WC1HAL UK h.joshi@ioe.ac.uk	**Patricia Schreuder** Norwegian Institute of Public Health Department for Project Support Kalfarveien 31 N-5018 Bergen Norway Patricia.Schreuder@fhi.no	