



MEETING REPORT

INTERCONNECT: A GLOBAL INITIATIVE ON GENE-ENVIRONMENT INTERACTION IN DIABETES AND OBESITY. *Funded by EU FP7 grant agreement 602068*

Monday 11 September 2017, European Association for the Study of Diabetes, Lisbon, Portugal

MEETING ABSTRACT

InterConnect seeks to optimise the use of existing data to enable new research into the causes of diabetes and obesity. Research questions can be difficult to address through the analysis of data from individual cohorts and instead often require cross-cohort analyses.

Physically bringing data together from cohort studies across the world is desirable because it allows great flexibility in the analyses and provides access to individual-level data, but it is constrained by governance, ethical and legal challenges. Alternative approaches also have challenges such as publication bias, analytical inflexibility, and major time commitment from local analysts.

The InterConnect approach enables cross-cohort analysis in a manner equivalent to a meta-analysis of harmonised individual level data but without any direct access to individual-level data. Analytical instructions are sent remotely and analysis is performed locally so all data stays at source and only results are shared; the analyst has the flexibility to refine and re-run analyses quickly, the local time burden is much less, and all analyses are conducted consistently across all participating studies.

www.interconnect-diabetes.eu

VISION FOR THE INTERCONNECT APPROACH (NICK WAREHAM)

Professor Nick Wareham, Co-ordinator of InterConnect and Chair of the meeting, welcomed the participants. The aim of the Symposium was to introduce the InterConnect approach and demonstrate its application. InterConnect aims to enable others by creating the foundations for cross-cohort analyses to enable research to move from explaining the differences in risk within populations to being able to explain differences between populations.

The presentation went onto discuss how to realise this vision. One option would be to create one large international cohort – but this is not really practical. Other options are to make better use of existing data and to be able better align prospective studies through use of comparable metrics. Prof Wareham noted that InterConnect is contributing to this latter option through the development of a measurement toolkit and that this would be presented by Matthew Pearce in the next presentation. He went onto discussion in more detail the steps required to make better use of existing data i.e the need to be able to:

- Find relevant studies globally
- Find out what data the studies have collected
- Find an appropriate way of bringing data together
- Find a way of interpreting different forms of data that are brought together

He then outlined the main elements of the InterConnect approach i.e. the study registry to facilitate data discovery, tools for data harmonisation and an analytical approach based on ‘taking the analysis to the data’.

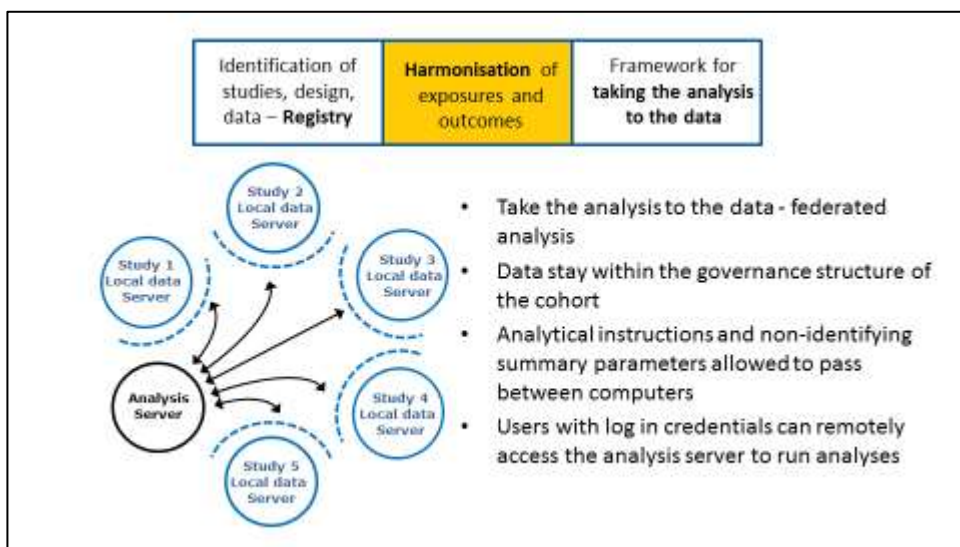
InterConnect: Live Study Registry

| Short Name | Name | Study Design | Actual number of participants recruited to the study | Country of residence |
|--------------|----------------------------------------------------------------------|--------------------------|------------------------------------------------------|----------------------|
| MEC | Medicine Contact Study | Prospective cohort study | 231 251 | United States |
| SWS | Southeastern Women's Survey | Prospective cohort study | 12 563 | United Kingdom |
| | Healthy Diet Study | Prospective cohort study | 3 620 | United States |
| ALSPAC | Avon Longitudinal Study of Parents and Children | Prospective cohort study | 14 547 | United Kingdom |
| AHS | Agricultural Health Study | Prospective cohort study | 49 655 | United States |
| AHC | Atherosclerosis Risk in Communities Study | Prospective cohort study | 15 752 | United States |
| DRBC | Diabetes National Birth Cohort | Prospective cohort study | 104 940 | Denmark |
| EPIC - Turin | European Prospective Investigation into Cancer and Nutrition - Turin | Prospective cohort study | 10 004 | Italy |
| MHS1 | Marsh Health Study 1 | Prospective cohort study | 121 730 | United States |
| NCHS1 | The Northern Manhattan Study | Prospective cohort study | 3 286 | United States |

Researchers need to know what studies are being conducted. To help with this discovery of data, InterConnect has created an online registry of studies. The approach has been to enable wide coverage of studies with a limited set of information that can largely be collated from information already in the public domain. This creates little burden for each individual study while enabling sign-posting of a large number of studies useful for cross-cohort analyses. The registry focuses on the key question of findability and

is now live, with marked geographical diversity from studies around the world. Professor Wareham encouraged participants to explore the searchable registry online www.studies.interconnect-diabetes.eu/studies and to submit their own studies to continue to increase its range.

Rather than physically pooling data – i.e. ‘taking the data to the analysis’ – the InterConnect approach is based on ‘taking the analysis to the data’. Analytical instructions are sent remotely and the analysis is performed locally so all data stays at source and only results are shared as non-identifiable summary parameters. In this way, it is possible to perform an analysis that is equivalent to a meta-analysis of harmonised individual level data and so the approach is called ‘federated meta-analysis’.



Federated meta-analysis provides a secure, scalable and sustainable approach to cross-cohort analyses. Participant data from contributing studies are held securely on geographically dispersed, study-based computers. Making the most of existing data requires harmonisation of related variables from different studies. High quality meta-data is required to assess the degree to which this is possible. Algorithms can then be developed to transform the data into a common format and can be stored and for re-use in the future.

InterConnect has taken forward a number of exemplar projects that address research questions of aetiological and public health interest. These exemplars help to understand and address the real-life issues that affect implementation and also ensure tools and infrastructure are aligned to their research use. The exemplars engage researchers, develop specific resources for data harmonisation derived from practical needs, and begin to establish a collaborative network for federated meta-analysis. Finally, Prof Wareham highlighted that results from two InterConnect research exemplars would be presented later in the programme: Physical activity in pregnancy and neonatal anthropometric outcomes, and fish intake and risk of type 2 diabetes.

RESOURCES FOR DATA HARMONISATION – THE DAPA TOOLKIT (MATTHEW PEARCE)



The Diet, Anthropometry and Physical Activity (DAPA) Measurement Toolkit is a free, web-based resource to assist researchers and public health or public end-users to identify methods for the assessment of diet, anthropometry and physical activity:

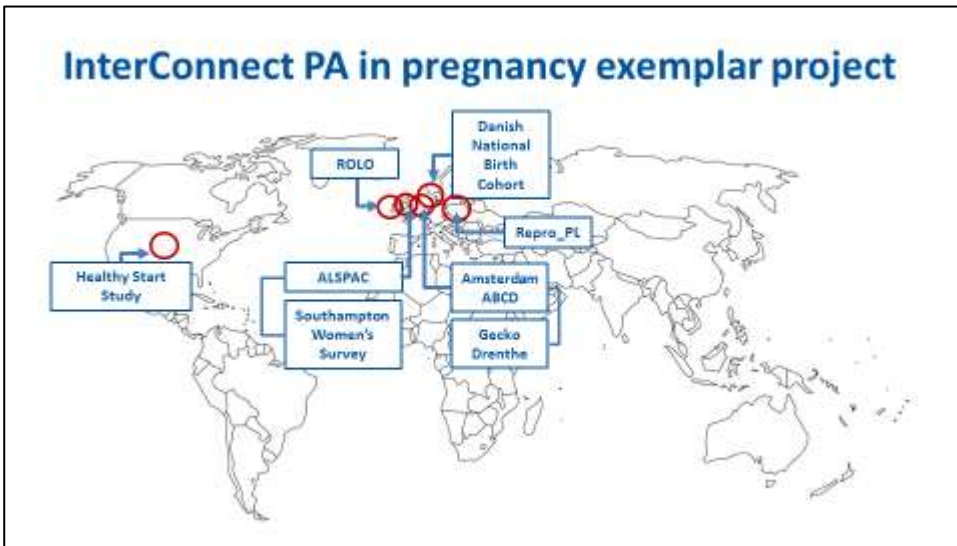
<http://dapa-toolkit.mrc.ac.uk/>

The toolkit is a rich resource for data harmonisation. It assists users when interpreting existing data and when selecting methods that

are fit for purpose when planning new studies. The DAPA toolkit provides an inventory of subjective and objective methods; dedicated harmonisation content and an instrument library. It also provides information on harmonisation concepts and case studies illustrating the harmonisation process. Dr Pearce explained how the toolkit facilitates both retrospective and prospective harmonisation. The toolkit aims to have the broader, more indirect, impact of promoting convergence of methods and compatibility of data by helping researchers to make the best decisions about which methods are most appropriate for interpretation of their studies. The team are working on: increasing the usability of the instrument library; a web-form for external researchers to upload their own instruments and resources; and an interactive map of relationships between different methods. A long-term goal is develop an interactive map so researchers will be able to see the relationships between different methods.

THE RELATIONSHIP BETWEEN MATERNAL PHYSICAL ACTIVITY IN PREGNANCY AND OFFSPRING BIRTH SIZE (SILVIA PASTORINO)

Dr Pastorino presented the results of the first InterConnect exemplar study, on behalf of the collaborative grouping of studies involved. Physical activity (PA) during pregnancy was identified as a significant area of research because of its potential for intervention to lower the risk of large offspring birth size (LGA and macrosomia), with positive benefits for pregnancy (obstetric) outcomes and longer-term obesity risks (for mother and child). Systematic reviews on the association between PA during pregnancy and offspring birth size exist but findings were highly heterogeneous, there is potential publication bias and many studies were unadjusted for confounding. In contrast, federated meta-analysis can reduce heterogeneity by allowing harmonisation of exposures and consistent consideration of confounding. It has also allowed investigation of modifying factors, the effect of different PA domains, the shape of the association and thresholds and timing of exposure to PA to be explored, and any potential publication bias is avoided.



A collaborative grouping of eight international studies formed around the research question 'What is the association between physical activity during pregnancy on neonatal adiposity?' Each study set up their local server, data harmonisation algorithms were developed and coded, and federated analyses have been completed. The analysis tested models before and after adjustment for potential confounders (maternal SES, age, parity, smoking, alcohol,

ethnicity). In this way, heterogeneity between studies was substantially reduced by consistent adjustment for confounders. In summary, the results indicated that:

- Leisure time moderate-vigorous physical activity during late, but not early, pregnancy has a small but significant inverse association with offspring birth size (BW, large BW, ponderal index)
- No association with higher risk of low birth weight (SGA)

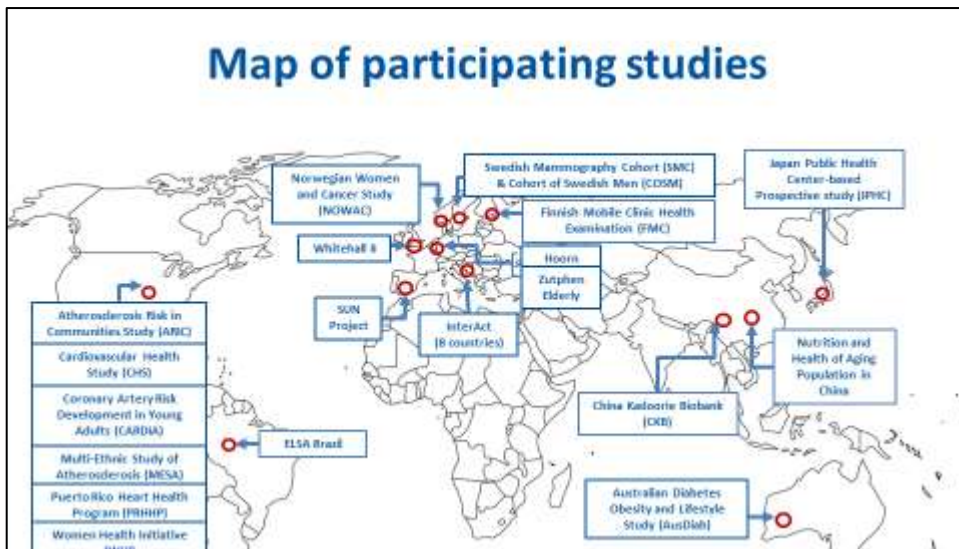
The scientific results for this exemplar have now been written up and submitted for publication.

THE RELATIONSHIP BETWEEN FISH INTAKE AND TYPE 2 DIABETES (NITA FOROUHI)

A number of studies are now collaborating to investigate the relationship between fish intake and new-onset type 2 diabetes which is an area where there is considerable scientific uncertainty. This exemplar also paves the way for further exemplar questions relating to other dietary factors. It is proposed that fish intake is likely to be beneficial for the prevention of type 2 diabetes, based on the benefits for cardiovascular health. Yet many studies find a null relationship between fish intake and the incidence of type 2 diabetes. A wide variety of conclusions have been drawn since 2012 and the debate continues. Dr Forouhi pointed to a number of significant research gaps:

- Systematic reviews analysed total fish and did not distinguish between types of fish (e.g. fatty fish, lean fish and shellfish) or cooking methods;
- Systematic reviews did not include unpublished results;
- High heterogeneity in meta-analyses might be caused by:
 - Different confounding structures of included studies
 - Different fish exposures (portions sizes varied across studies)

The InterConnect approach helps to advance this research question, as it enables use of individual participant meta-analysis without physical pooling of data, heterogeneity can be reduced by including the same types of confounders and harmonising exposures and outcome to a common format and by including studies on the association between fish and type 2 diabetes that have not yet been published.



Twenty cohorts from across the world are now collaborating to investigate this research question. The data have been harmonised across the cohort and a wide range of confounders were taken into consideration.

Dr Forouhi went on to present the initial results. In summary, there was no significant association between either total fish, or types of fish (fatty and lean) or cooking method (fried fish) and type 2 diabetes. The heterogeneity was, however high, close to 70%. A

more detailed approach was then taken, accounting for confounders, but again, the overall relationship between total fish intake and risk of type 2 diabetes remained null. Conclusions on the effect of geographical location were that:

- Heterogeneity was reduced when results stratified by geographical location ($I^2=35%$ for US, $I^2=54%$ among Eastern countries)
- Higher T2D risk among US studies (HR for 120g/week=1.015, $p=0.01$)
- Null associations for European countries
- Tendency for lower risk among Asian countries but not Australia

In terms of next steps in this exemplar, more studies will be added to the next stage of federated meta-analysis:

- China Kadoorie Biobank (China), $N>500,000$ m/w
- Shanghai Women's Health Study (China), $N>70,000$ w
- SUN Project (Spain), $N=22,340$ m/w
- Finnish Mobile Clinic Health Examination (Finland), $N=4,304$ m/w

Future analysis will be more nuanced, including the stratified analysis by region and types of fish. Sensitivity analyses will also be undertaken, using other models including waist circumference, family history of diabetes and fish oil supplements and test interaction for sex and BMI.

DISCUSSION OF THE CHANGING LANDSCAPE FOR CROSS-COHORT ANALYSIS (NICK WAREHAM)

These two exemplars have demonstrated how the InterConnect approach can address research questions of aetiological and public health importance. Additional diet-related research questions are also in progress and the grouping involved in the first exemplar are now considering new research questions such as the association between birth weight and childhood central fat deposition. However, driving cross-cohort work on a larger scale requires many actors, not only researchers:

- Stakeholders who are users of research evidence – create pull;
- Funders – change infrastructure, incentives for re-use of data. They can invest in discoverability and promote the development of accessible study meta-data. Funder and journal recognition of data sharing by researchers is also essential.

Professor Wareham drew attention to those situations where federated meta-analysis might be useful, particularly when analysis in published literature is inadequate: Publication bias is very difficult to overcome analytically, and pooled analysis of data is sometimes not possible such as when it is potentially identifiable e.g. geospatial data. He was pleased to note that a wider pull for the approach may be beginning e.g. the new H2020 LifeCycle project plans to use a DataSHIELD platform for federated meta-analysis of studies in the EU Child Cohort Network. After some general discussion about the InterConnect initiative, Prof Wareham drew the meeting to a close, thanking participants for attending and for their welcome contributions to the meeting.

SPEAKER BIOGRAPHIES



Dr Nita Forouhi, Group Leader, Nutritional Epidemiology Programme, MRC Epidemiology Unit, University of Cambridge, UK. Dr Forouhi leads the Nutritional Epidemiology Programme, which aims to understand the relationship between diet, nutrition and the risk of diabetes, obesity and related disorders. Dr Forouhi is also an Honorary Consultant Public Health Physician with Public Health England. She is the Chair of the area multidisciplinary Managed Care Network for Diabetes, a member of Diabetes UK Research Committee and Associate Editor of *Diabetic Medicine*.



Dr Silvia Pastorino, Career Development Fellow, MRC Epidemiology Unit, University of Cambridge, UK. Dr Pastorino is involved in the coordination and research development of international consortia as part of the InterConnect project. Dr Pastorino obtained her PhD in Epidemiology and Public Health at University College London investigating dietary and anthropometric determinants of type 2 diabetes.



Dr Matthew Pearce, Research Associate, MRC Epidemiology Unit, University of Cambridge, UK. Matthew studied for an MSc in Nutrition, Physical Activity and Public Health at the University of Bristol before completing his PhD at the University of Edinburgh in 2015. This research focused on the use of combined accelerometer, GPS and diary methods to understand the context and intensity of children's physical activity, particularly that which occurs outdoors. Following work as an analyst within the banking industry, Matthew was an Associate Research Fellow at the University of Exeter before joining the Unit. He is working on the Diet, Anthropometry and Physical Activity (DAPA) Measurement Toolkit as part of the InterConnect project.



Professor Nick Wareham, Director of the MRC Epidemiology Unit and co-Director of the Institute of Metabolic Science. His principal research interests are the aetiology and prevention of obesity and diabetes. He also leads the Aetiology of Diabetes and Related Metabolic Disorders Programme, is the Director of the Centre for Diet and Activity Research (CEDAR) and the Coordinator of the InterConnect project. Nick qualified from St Thomas' Hospital Medical School, London, and trained in epidemiology at the London School of Hygiene and Tropical Medicine, London. He then worked at Harvard University, Boston, USA before joining the University of Cambridge where he became a Wellcome Trust Senior Fellow in Clinical Science. Now he is a Professor and an Honorary Consultant at Addenbrooke's Hospital, Cambridge. He is also the Clinical Research Lead for Division 2 (diabetes, stroke, cardiovascular disease, metabolic and endocrine disorders, and renal disorders) of the NIHR Clinical Research Network: Eastern.