Global data for diabetes and obesity research

MEETING REPORT

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

Friday 10 October 2014, Club of the University Foundation, Brussels, Belgium

The InterConnect project (www.interconnect-diabetes.eu) was set up by the European Union following their report on establishing a global initiative to study gene-lifestyle interaction on diabetes and obesity. It aims to enable a new way of sharing data from individual studies around the world to enhance our ability to investigate why risk varies significantly between different populations.

A key aspect of the InterConnect approach is that data stays within the governance arrangements of the organization that collected it but use of data from many different studies around the world is optimised. It builds on the work of a number of research groups that have developed open-source tools for data harmonisation and federated analysis of data and will apply and adapt these tools to develop a scalable and sustainable platform to enable population research into the causes of diabetes and obesity. **The approach is egalitarian and democratic as it enables all those with responsibilities for studies to play a central role rather than simply providing data for others.**

A significant challenge is to get scientists, research funders and stakeholders to engage with InterConnect as an initiative. We are engaging scientists by developing exemplar projects that illustrate how the approach can make novel and otherwise challenging research questions tractable.

The particular focus of this meeting in Brussels on 10 October 2014 was to develop a shared vision with funders and stakeholders of the challenges of current data sharing models and how InterConnect can change this paradigm. The role that all groups can play in incentivising and encouraging data sharing and the potential to create a forum for discussion of future collaborative research enabled by InterConnect was also addressed. The meeting comprised presentations with table-based discussion facilitated by members of the InterConnect team.

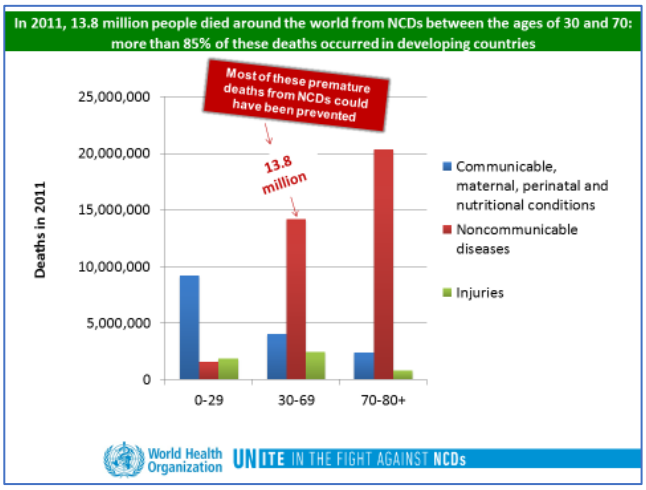
SESSION 1: SETTING THE SCENE

This session linked to the international conference hosted by the European Commission in February 2012 on *'Diabesity – A worldwide challenge: towards a global initiative on gene-environment interactions in diabetes/obesity in specific populations'* and showed how the vision from that meeting is being taken forward by InterConnect. It described the scientific opportunity that arises from enabling research to move from explaining differences in risk of diabetes and obesity *within* populations to being able to explain differences in risk *between* populations and how data sharing is the key approach to achieving this transition.

Professor Nick Wareham, co-ordinator of InterConnect and Chair of the meeting, welcomed the participants to the meeting and introduced Dr Nick Banatvala, Senior Adviser to the Assistant Director General, Non-Communicable Diseases and Mental Health, at the World Health Organization.

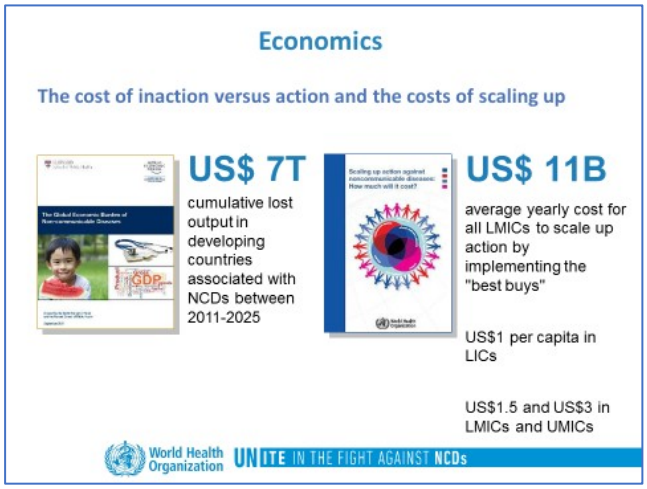
PERSPECTIVES FROM THE WORLD HEALTH ORGANIZATION: NICK BANATVALA

Non-communicable diseases (NCDs) represent one of the world's major development challenges, both in terms of the great human suffering they cause, as well as the immense harm they inflict on the socio-economic fabric of many countries, particularly those of the world's poorest peoples. In 2011, 13.8 million people worldwide died from NCDs between the ages of 30 and 70: more than 85% of these deaths occurred in developing countries. Today, more than 8 million people die before the age of 60 in developing countries from NCDs. Diabetes is a major contributor to this burden.



The rise of NCDs represents a global crisis and we are justified in our use of the word 'epidemic' to describe it. It is a priority - both global and political – to address the crisis surrounding diabetes and chronic disease more generally. Understanding the causes of diabetes is crucial to this. The World Health Organization (WHO) as a global health organization recognises that partnership is vital, working together to respond better to public health challenges. In this respect, Dr Banatvala welcomed InterConnect as representing an exciting new global initiative.

NCDs represent one of the world's major development challenges, both in terms of levels of human suffering, and their subsequent socioeconomic impact. Deaths due to NCDs in the 30-60 age group are particularly striking. We all strive for a peaceful death yet we are witnessing unnecessary suffering and premature deaths - many of which could have been prevented. The global incidence of cancer, stroke and heart disease is increasing whereas some communicable diseases are decreasing so we need a shift in perspective onto these issues.



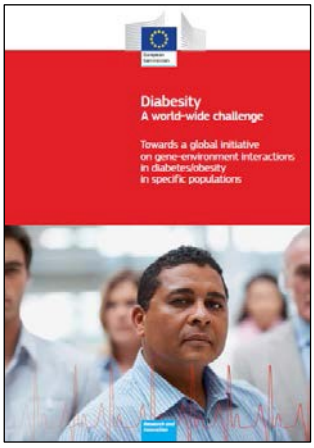
Poverty contributes to NCDs and NCDs cause poverty. The cost of inaction far outweighs the cost of action between now and 2025. Diabetes is a major global issue, as is obesity, with staggering figures only set to rise. Diabetes will continue to rise, especially in Africa, but with an increase in all countries. Underweight and obesity currently coexist in middle and low-income countries. Responding to the challenges of NCDs and diabetes is vital, and these are now at the top of the global health agenda. In 2011, the United Nations (UN) General Assembly highlighted the importance of multi-sectorial action, with roles for the UN, academia and civil society. One critical action is the WHO Global NCD Action Plan 2013-2020. It's vision is a world free of the avoidable burden of NCDs, with the goal to reduce the preventable and avoidable burden of morbidity, mortality and

disability due to NCDs by means of collaboration and cooperation at national, regional and global levels. The Action Plan has six objectives and national capacity for high quality research is one of them.

Diabetes and obesity will doubtless be included in forthcoming global monitoring exercises and the millennium development goals. Political pressure continues, with a follow-up meeting in New York in 2014, at which the 2014 UN Outcome Document on NCDs (Resolution A/RES/68/300) saw governments commit to concrete action and map out concrete national commitments between 2014 and 2018.

The World Health Organization also has a prioritised research agenda for NCDs, with specific attention being given to diabetes, obesity and genetics. **Dr Banatvala therefore concluded by reiterating that the InterConnect initiative is valued highly by the World Health Organization as an opportunity to better explain the causes of diabetes and to work more effectively at a global level.**

THE EU DIABESITY CONFERENCE 2012: NICK WAREHAM



The 2012 EU Diabesity conference began the process of driving towards a global initiative to tackle the crisis**.** The report http://ec.europa.eu/research/health/pdf/diabesity-conference-report-022012_en.pdf concluded that **we need research into individual and societal approaches to the prevention of obesity, diabetes and related metabolic disorders, particularly around prevention and intervention in order to mitigate risk. The need for further research into understanding differences in individual and population risk, and to develop better ways of working together to promote better understanding of risk were also highlighted**. Professor Wareham then went on to introduce Dr Karim Berkouk, Deputy Head of Unit at DG Research and Innovation at the European Commission (EC). The EC is committed to tackling diabetes intervention research through its membership in the Global Alliance for Chronic Disease (GACD) (www.GACD.org).

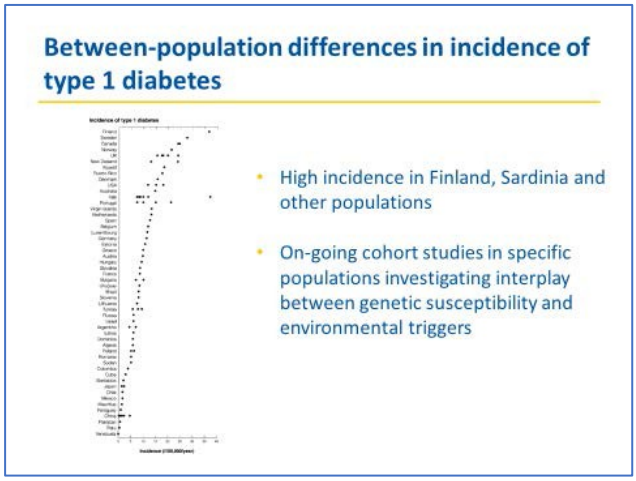
GLOBAL ALLIANCE FOR CHRONIC DISEASE AS A VEHICLE FOR INTERVENTION RESEARCH: KARIM BERKOUK

Dr Berkouk began by saying that chronic disease and ageing are major challenges in the European context, so the European Commission (EC) welcomes this opportunity to support InterConnect. Ageing is a global issue and NCDs are a major problem, since they require treatment for the life of the patient. It costs in the region of 50,000 euro per year to cover the treatment costs for one patient with chronic diseases such as cancer, which is clearly unsustainable. The EC has traditionally adopted a bilateral approach to this situation, but this has transitioned to a global level due to the effectiveness of multilateral agreements, which is preferable given the scale of the challenge. The GACD is a good model to draw on in this context – this global approach is working and therefore should be scaled up.

The primary focus of the GACD diabetes call is to provide a vehicle for intervention research and is therefore complementary to InterConnect which will enable understanding of the causes of the disease. More work is needed to address a contextual, individualised approach – what works, for whom and when. The aim is to create a global network of researchers in different global regions in order to meet the challenge of NCDs.

Dr Berkouk went onto emphasise that knowledge should not remain in the research arena but must be implemented. Patients and care providers must be the beneficiaries of research efforts, and research must be implemented once interventions prove to be successful.

UNDERSTANDING DIFFERENCES IN RISK OF DIABETES AND OBESITY BETWEEN POPULATIONS: NICK WAREHAM



InterConnect is about understanding the causes of diabetes and obesity, specifically what explains the highly significant differences in risk **between** populations. Nick Wareham first described research into the incidence of type 1 diabetes. Finland has one of the highest levels of incidence, as does Sardinia; however, there are marked differences in the complex interplay between genetic and environmental factors. A different pattern exists for type 2 diabetes, with very high prevalence in some counties but not in others. Some studies have tried to take this further, looking for possible explanations for differences in risk. In 1962, Professor James Neel developed a hypothesis based on 'thrifty genes' which enable individuals to efficiently collect and process food to deposit fat during periods of food abundance and, in 1992, the 'thrifty phenotype' hypothesis in which reduced fetal growth due to adaptations made by the fetus in an environment limited in its supply of nutrients is strongly associated with a number of chronic conditions later in life, was proposed.

New studies are now trying to understand individual risk **within** populations in more detail. The InterAct project www.inter-act.eu funded under the EC Framework 6 Programme, is a study of half a million people, including a cohort of 12,403 with type 2 diabetes. InterAct is analysing the effects of foods that might be associated with the incidence of type 2 diabetes, such as fizzy drinks and processed meat, alongside those foods that might be protective, such as fish, fruit, vegetables and certain dairy products; it is also looking at physical activity and how this might be protective. Given the project's scale and its robust data, InterAct is an important study in the field. However, even with a study of this size, no evidence of interaction with individual genes of known variants has been found and therefore global analyses across multiple studies are required if we are to further research on gene-environment interaction.

Professor Wareham went on to explain how the research community now needs to explore how to study the large differences in risk that exist **between** populations - and how InterConnect is rising to this challenge. Globally, between-population differences in both genetics and also lifestyle or environmental factors are considerably larger than differences within-populations. Unless we think globally we cannot tackle this problem.

To realise the vision of bringing data together to allow the study of between-population differences in risk, InterConnect aims to help researchers 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.

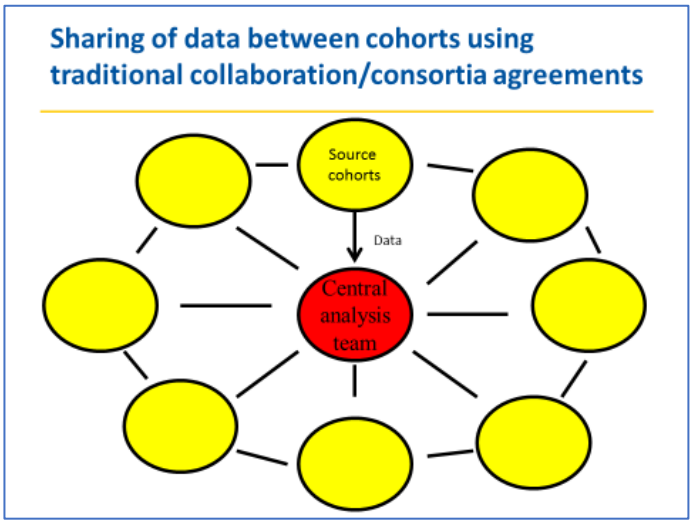
In this way, InterConnect aims to create the foundation for a sustainable, global network for diabetes and obesity population research that enables research to move from explaining differences in risk within populations to being able to explain the major variations in risk between populations.

SESSION 2: CHALLENGES OF CURRENT DATA SHARING MODELS

Current models of data sharing, both those aimed at making the best use of existing data and those that seek to optimise the collection of future data, present challenges. This interactive session worked through different models using facilitated discussion at each table to hear the perspectives from different types of stakeholders.

Professor Wareham outlined the structure for Session 2, in which four alternative models were presented on how data is currently shared, and it might be shared in the future. Participants were invited to think through these different models, the possible benefits and difficulties of each from different perspectives. Ultimately, groups were to try to imagine a future world in which we are trying to connect multiple studies at a global level.

Model 1: Sharing Data between Cohorts



The old model consists of physically sharing data, typically involving collaboration between a number of cohorts and a central analytical team where data flows between institutions. Participants were asked to consider to what extent this model would be feasible if researchers tried to scale it up globally? The InterAct project was given as an exemplar of this model. It was a product of major investment in early 1990s which funded a number of cohort studies in different countries. While it would be ideal to think that funders around the world will set up similarly standardised ways of collecting data, this is unlikely on the scale required. Moreover, this model creates a distinction between the people who collect the data and those who interpret and analyse it.

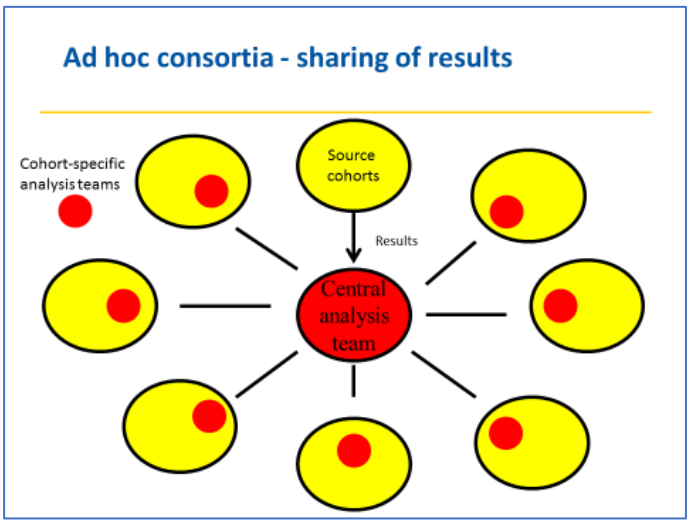
Key roundtable discussion points:

- This model enables physical sharing of individual level data, and in general it is better to conduct meta-analyses between studies with this type of data.
- The model is problematic in terms of exchange of data and the regulatory issues relating to cross-border transfer.
- This model requires well established collaborative networks between partners, the development of which is a lengthy process and requires trust.
- A one-off effort is required by funders to establish comparable studies and further agreement between institutions may become difficult as the project changes over time.
- Data transfer problems and diversity of attitudes can be limiting; bringing in a global perspective will add significantly to the complexity and make comparative global studies virtually impossible.

SUMMARY

- There is a considerable transactional burden – many institutions cannot cope
- The burden will increase exponentially as number of partners in consortia increases
- It is difficult to control passage of data and use beyond the original intention
- If centralised around a sole analytical centre, resentment will arise about imbalance of opportunities to lead analyses as opposed to contributing data

Model 2: Ad Hoc Consortia – Sharing of Results



This model has been shown to work especially well for genetic data. Typically, each cohort has its own analytical team and does its own analysis; results are then submitted by each cohort and collated centrally to complete the analysis.

Key roundtable discussion points:

- Some of the ethical issues are eased here as the data remains in the control of the researcher who collected it.
- Each cohort has to have analytical capacity otherwise it cannot contribute its results to the central analysis.
- Even if an organization is well resourced, there will be a massive number of requests for data and the investigator could find themselves servicing other people's research and not their own. This would also be of concern to the funders of their research.
- The quality of the research conducted by each cohort could be variable as the central analysis team cannot control or standardise other cohorts' work within the project. This is an issue for both researchers and funders.
- A full global comparison requires analysis of data across research centres / cohorts. When researchers meta-analyse results rather than data, they may miss important details when analysing across populations.
- Sharing results has its limitations, particularly if the cohort is drawn from a heterogeneous sample of the population. Researchers are now more frequently looking at gene-environment interaction in terms of race and ethnicity, and this approach is more limited to homogenous groups. A new approach is clearly warranted if researchers want to undertake more complex analyses.

SUMMARY

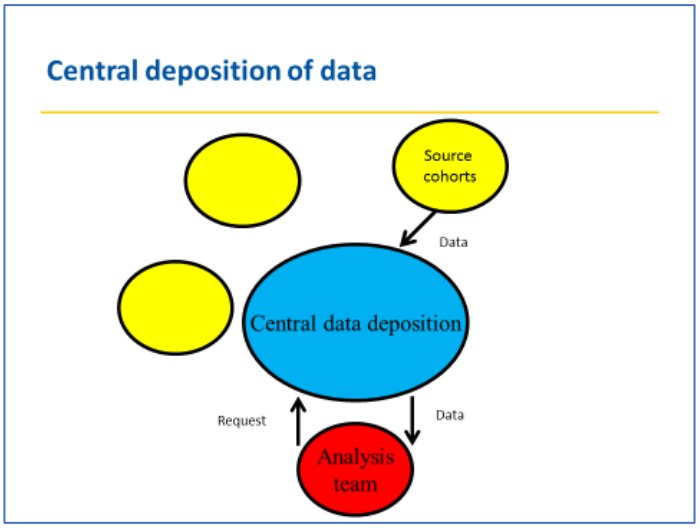
- Ad hoc consortia work well for genetic analyses, allowing sharing of results without administrative or organizational complexity
- There are limits to meta-analysing interaction terms from individual studies
- The difficulties of data harmonisation are often given limited attention
- The analysis is potentially missing major between-cohort variation
- The analytical effort is decentralised to individual studies who spend a massive amount of time servicing the work of others

Model 3: Central Deposition of Data

In this model, data are deposited into a central repository. The analytical team puts in a request and can then undertake analyses.

Key roundtable discussion points:

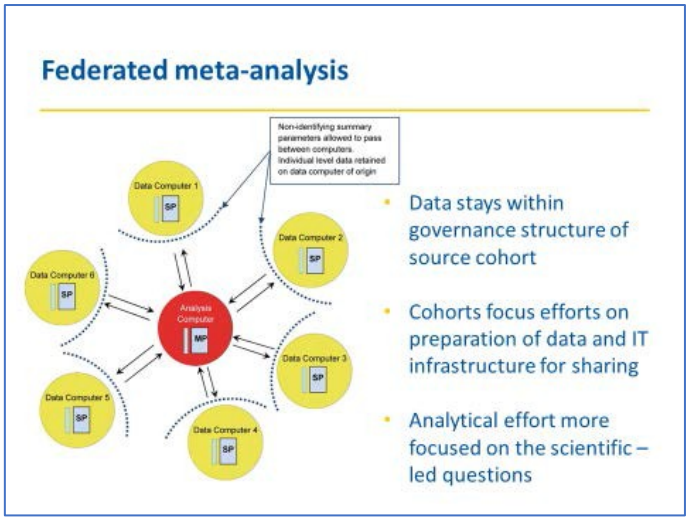
- This model requires a very well-standardised approach from the outset and if this is not the case, achieving reliable results can be a major challenge.
- This is a more democratic approach and provides greater opportunity to a wide range of researchers to access the data.
- This model does have advantages if a cohort does not have analytical capability, but governance issues still exist around who holds and owns data and how the data is centralised; there are also issues around potential duplication of work and confidentiality.
- This model is not sustainable and access decisions require delegated authority, which would be a significant challenge on a global scale.



SUMMARY

- The approach works within some countries for some forms of data
- The likelihood of success for between-country collaboration is low
- It is unlikely to work for more complex forms of data
- There are major governance, ethical and legal challenges
- It is difficult to mandate for historical data

Model 4: Federated Meta-Analysis



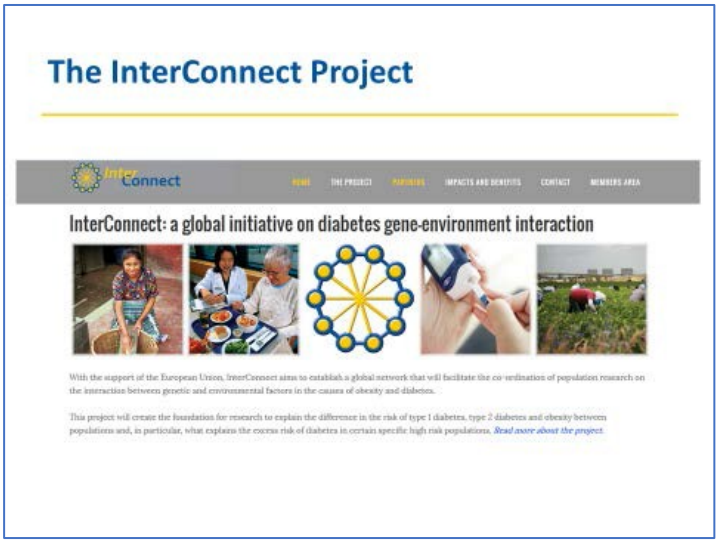
There are different issues at work with this final model, which was introduced as a prelude to the following session on changing the paradigm. The model is not free from problems, but may offer a number of solutions.

The data stays behind the firewall of wherever the source cohort is located; raw data does not move. Cohorts focus their effort on preparing the data and providing access for data-sharing purposes. The analytical effort is then more focused on science-based questions. This 'federated meta-analysis' approach, if feasible, can facilitate a democratic system where all partners can equally drive analyses. This model avoids some of the organizational complexity as extensive contractual agreements between cohorts are not needed; rather, cohorts maintain control of their own data while being allowed access to a wider spectrum of data. The afternoon session will expand on this vision of a changed paradigm, and how together we can work towards achieving this.

SESSION 3: VISION OF A CHANGED PARADIGM

This session comprised an overview of InterConnect and how it aims to change the way in which data are used. The three main elements of the platform were described: a study registry that enables scientists to identify the full range of available resources; tools and processes for the harmonisation of data; and a secure, federated network that provides a mechanism for meta-analysis of individual participant data while keeping the data within the local organization that collected it.

INTRODUCTION TO INTERCONNECT: NICK WAREHAM



InterConnect provides a new approach to data sharing that is secure, scalable and sustainable. It builds on the work of a number of research groups, particularly Maelstrom Research and the EU funded BioSHaRE project, that have developed open source tools to catalogue studies, support data harmonisation and enable federated meta-analysis. InterConnect provides a bridge between these tools and implementation by the diabetes research community. It is vital that researchers, funders and stakeholders are involved in the development of the InterConnect approach at this early stage. Professor Wareham introduced the members of the consortium, highlighting two of the InterConnect Investigators, Dr Isabel Fortier and Professor Ronald Stolk, involved with Maelstrom Research and the FP7 funded BioSHaRE project**.** He went on to describe the major themes within InterConnect that are developing resources to support wider research i.e. the development of a study registry, tools for method and data harmonisation and consideration of the methodological and governance issues related to data sharing.

STUDY REGISTRY: MATTHIAS SCHULZE

Professor Schulze explained that a major part of InterConnect is the development of a study registry. This is necessary for researchers to be able to find out:

- What resources are available globally
- What study design was employed
- What populations were recruited
- Whether samples were stored
- What data is available

The first phase of the registry is taking a 'broad and shallow approach' and is focused on gathering simple but useful information that can largely be collected from information already in public domain such as general information (study name, contact persons, web link), study design, ethnicity and race, the sampling frames, recruitment information and basic participant characteristics. This approach creates little burden for individual studies while enabling a sign-posting of large number of useful studies. The second, more detailed phase of the registry development will involve information to be collected directly from studies and will incorporate metadata about available data such as data sources and categories of available data (e.g. health, socio-demographic, lifestyle, physiological, biochemical, genotype information).

How to create the commitment of studies to provide information for the phase 2 development is an important issue along with promoting its use. It is hoped that the two phase approach will be useful in this regard, stimulating interest via the breadth of studies included initially and populated by collating publicly available information. The registry is part of a larger platform to support data harmonisation and federated meta-analysis which is likely to encourage involvement from studies. Other, longer term considerations included keeping the registry up to date,

both in terms of new studies and also new data collection events within existing studies, and sustainability of the infrastructure.

SUMMARY

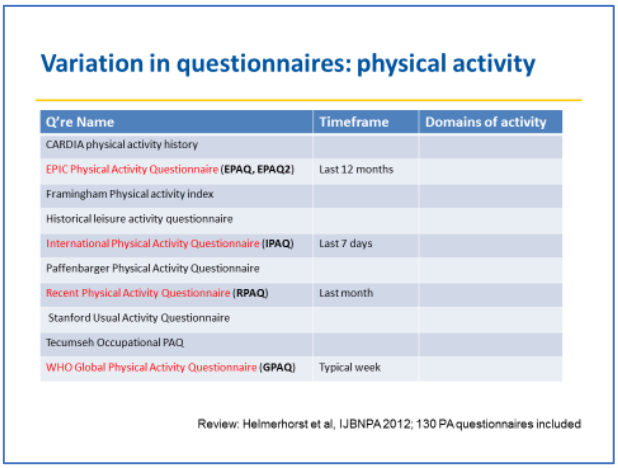
- The study registry aims to ease the current challenge of study 'findability'
- There is minimal work for cohorts - InterConnect can populate the first phase of the study registry for researchers to demonstrate utility
- The registry is part of a wider platform for data harmonisation and federated analysis which will encourage commitment and use

DATA HARMONISATION: NITA FOROUHI

Dr Forouhi explained the next stage of the process, data harmonisation which:

- Is about optimising data from single studies for re-use in combined analyses across multiple studies
- Achieves standardisation of data across different studies to a common format to maximise the data value from each study
- Involves re-coding or modifying variables so that they are comparable across research studies
- Enables the synthesis of primary data from studies

The context for data harmonisation is that data are collected for variables of 'exposures' and 'outcomes'. Exposures and outcomes are often assessed using different methods, as appropriate to the specific question or context of the study, and hence the data may need to be harmonised into a common format before analysis across studies is possible. Retrospective harmonisation of data can be a viable option to make best use of existing data. The InterConnect registry will list available studies and related meta-data to assess the harmonisation potential; the analysis platform (see presentation from Dr Fortier below) then enables algorithms to be applied that transform the data into the common format prior to analysis.



Using the example of measuring physical activity, it becomes clear that a whole host of options exist in terms of choices of measurement from objective methods such as accelerometry to subjective methods such as questionnaires. Hence, it is difficult to harmonise data on a large scale given the wide divergence in variables, even when studying the data from the same type of measurement, such as physical activity questionnaires with different time-frame, domains of activity, and number of questions. While researchers can harmonise data to a certain extent, they often have to exclude specific studies since it is not always possible to retrospectively harmonise data. InterConnect will, therefore, also develop approaches for prospective data harmonisation. This will enable greater comparability of future studies, defining and agreeing the optimum measures and procedures across studies up-front. This process will be aided by toolkits that signpost researchers to methods that are fit for purpose and feasible in specific settings for both self-reported and objective methods.

SUMMARY

- Retrospective harmonisation makes the best use of existing data and is facilitated by the InterConnect registry and federated analysis platform.
- There are limits to the feasibility of retrospective data harmonisation and InterConnect will provide a framework to enable greater comparability of future studies.

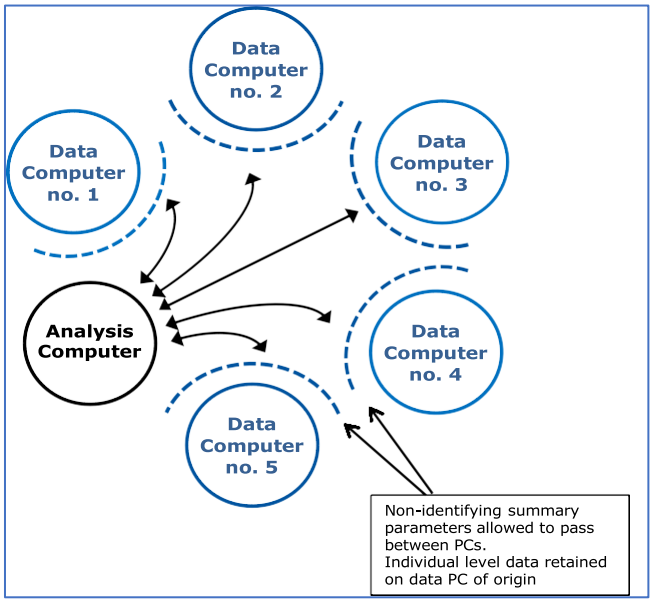
FEDERATED ANALYSIS: ISABEL FORTIER

Dr Fortier outlined how InterConnect provides a new approach to data sharing which is secure, scalable and sustainable. It builds on the work of a number of research groups that have developed open-source tools for retrospective harmonisation and analysis of data and will apply and adapt these tools to develop a global collaborative network for diabetes research. A fundamental aspect of the approach is a *federated* process. Individual participant data from contributing studies are held securely on geographically-dispersed, study-based computers; analytical commands are sent as blocks of code from a computer within the network which requests each computer to undertake an analysis and return non-identifiable summary statistics (i.e. results, not data). Analyses are performed locally so all data stays at source, within the governance structure and control of the originating study.

In order to move towards a federated process, we need to:

- Develop a collaborative framework (Investigators open to collaboration, ready to invest time and resources)
- Understand input data (study designs; what and how data was collected; quality of study-specific data)
- Ensure rigour (systematic harmonization process and quality control)
- Be transparent (document how the harmonized variables are created to permit reproducibility and long-term usage)
- Facilitate access (develop infrastructures permitting secure and efficient access to data).

InterConnect builds on the work of several research groups that have developed an integrated support platform for retrospective harmonisation and federated analysis of data; these have been tested in the FP7-funded BioSHaRe project (www.bioshare.eu) led by Professor Ronald Stolk, and made available as an open source toolkit by Maelstrom Research (www.maelstrom-research.org).



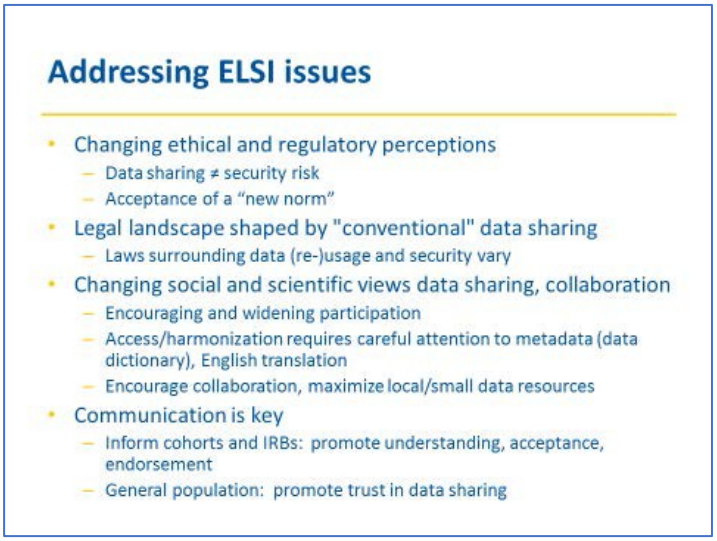
BioSHaRE's Healthy Obese Project comprises 10 studies with over 200,000 participants across 7 countries. Within firewalls of each of these organizations the required software was downloaded and data was imported onto a local server at their site. Research centres kept full control of access to the data and could also remove data if necessary. All the analysis was then done remotely. Dr Fortier emphasised that the first step before conducting the analysis was to understand the data and that in order to do that it was vital that the project construct a central catalogue, a truly searchable registry with high quality metadata. The tools for federated analysis continue to be developed and InterConnect is the ideal vehicle to promote uptake and use by diabetes researchers. InterConnect is the means by which we can connect the tools to the researchers and thereby connect data from different studies to meet the needs of population research on diabetes and obesity.

SUMMARY

- The federated process means that participant data from contributing studies is held securely on geographically dispersed study-based computers
- Analyses are requested remotely and performed locally, so all data stays at source under the governance structure and control of the originating study.
- Analyses equivalent to individual participant data meta-analysis are enabled on harmonised data but only non-identifiable results and not data are returned to the analysis computer.

ETHICAL, LEGAL AND SOCIAL ISSUES (ELSI): RONALD STOLK

Professor Ronald Stolk is the co-ordinator of the aforementioned FP7-funded BioSHaRE project, which has opened the way for InterConnect. He reiterated that federated data analysis is a way of bringing the statistics to the data not vice versa. As in the previous presentation, he emphasised that individual data stays within the original cohort and is never physically shared; analyses are therefore considered to be equivalent to using anonymous data. It is impossible to externally access personal data, as it never moves beyond cohort firewalls, alongside other safeguards such as restricted coding. This methodology has been proven safe after decades of investigation. The ethical issues are therefore no different to the usual ethical issues associated with traditional collaborative projects e.g. compliance with the terms of consent under which the data was originally collected.



Professor Stolk went on to add that, in terms of the current legal landscape, there is little legislation relating to data sharing. However, there is an International Code of Conduct for Genomic and Health-Related Data Sharing which promotes access to shared data, knowledge, and resources. Ultimately, the Code will hopefully serve to promote data sharing and to sanction misuse. However, change in ethical and regulatory perceptions is required to transform the ways in which the scientific community and wider society view data sharing – it need not be judged solely as a security risk but can be viewed as a 'new norm'.

InterConnect therefore has the potential to take open access publishing one significant step further – so that data becomes truly accessible, but anonymity is retained. Participants were invited to attend a forthcoming meeting in order to continue discussion of this issue, the BioSHaRE Tools roll-out meeting 28 July Milan, Italy.

SUMMARY

- Federated data analysis deals with the privacy issue that confounds some models of data sharing
- The ethical issues are no different to those associated with traditional collaborative project
- A change in ethical and regulatory perceptions is needed to create a new norm around data sharing, along with a way of recognising researchers for participation in data sharing

SESSION 4: NEXT STEPS

Session 4 focused on engagement and next steps. It covered what it is about InterConnect that will encourage scientists to share their data. It went on to focus on what InterConnect offers to funders of research, with an interest in maximising the value of their investment, as well as the wide range of stakeholder organizations with interest in the policy, social and economic benefits of diabetes research that would be enabled by more effective data sharing. The session closed with a vision of the role that all groups can play in incentivising data sharing and steering the InterConnect initiative, a summary of the conclusions of the meeting and specific actions.

FUNDER AND STAKEHOLDER ENGAGEMENT, CONCLUSIONS AND ACTIONS: NICK WAREHAM

Professor Wareham opened this final session by emphasising that this ambitious new paradigm will only become a reality with support from research funders and users of research – the European Commission and individual funders cannot do this alone. A vital part in this process is also, clearly, engaging with researchers, and ensuring that they are involved in the process from the outset. Professor Ruth Loos took up this latter issue in her presentation.

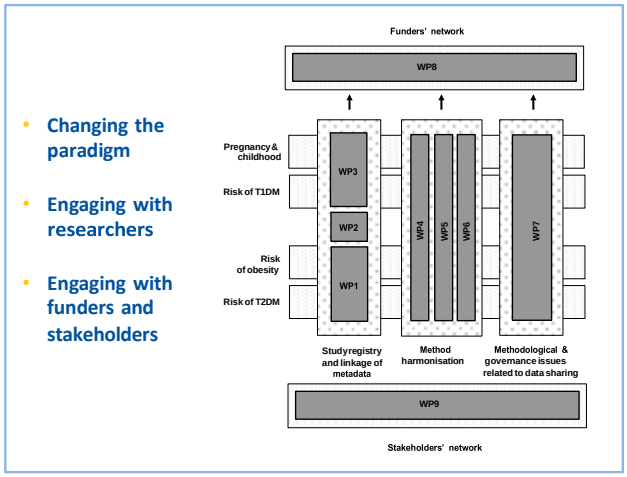
RESEARCHER ENGAGEMENT: RUTH LOOS

Dr Loos began by outlining the current situation regarding researchers' perspectives on data sharing. In the United States, researchers who hold projects funded by the National Institutes of Health (NIH) are obliged to deposit their data with the research community (using dbGap). Other researchers can then apply and download the data for their own research. A more collaborative approach is reached when researchers decide to work together and share data in the context of a consortium. A minority of consortia will share individual-level genotype and phenotype data, but this typically entails a significant administrative burden as contracts and agreements will need to be negotiated before this data can be shared. The majority of consortia share summary statistics of association analyses for central meta-analyses. This approach has been very successful, but has limitations, as described in session 2. The main limitation is that the heterogeneity is predominantly captured within each study, whereas the diversity between studies (and thus statistical power) is substantially reduced.

Researchers are increasingly realising that projects require large sample sizes to address more complex questions, for example on how lifestyle influences the genetic susceptibility to diabetes. Such gene-lifestyle interaction questions are hard to address in the typical consortia that combine summary statistics. InterConnect, however, aims to enable such research by using individual-level data from a large numbers of studies, yet without having to physically gather this data in one central data hub.

The federated approach InterConnect proposes is new to many scientists. Therefore, when considering the opportunity provided by federated meta-analysis, researchers will need to be convinced of a number of things:

- **Scientific question:** Innate need for federated analyses
- **Infrastructure:** Practically possible
- **Demonstration of feasibility:** Are others participating?
- **Return of investment:** Opportunity to lead projects, authorship



Even if it has been demonstrated as being technically possible, researchers need to be reassured that the approach will be useful. They have so much invested in their own projects that they are understandably concerned about sharing their data. They will need to be convinced that there is an innate need for federated analysis (i.e. that other approaches, such as using summary statistics, are not feasible), that it is practically possible and that others are equally committed to participation, so they are not 'going alone'. They will need to see a return on the investment they make in getting set up for this approach. Researchers need to be aware that this could be an opportunity to lead

projects and not simply contribute data for others to use.

The InterConnect team's impression is that researchers will engage with an infrastructure when they see a scientific purpose. For this reason InterConnect is developing a number of initial **exemplar projects** to illustrate how the approach can make novel and otherwise challenging research questions tractable. These currently comprise:

- What period in gestation is critical for the effect of gestational weight gain on foetal / neonatal adiposity?
- What is the effect of saturated fat on the incidence of type 2 diabetes?
- What is the effect of maternal supplementation with vitamin D during pregnancy and early life on the risk of type 1 diabetes?
- What is the effect of genes by physical activity interaction on body mass index by using objective measures of physical activity?

CONCLUSIONS AND ACTIONS: NICK WAREHAM

To conclude, Professor Wareham drew an analogy with open access publishing. This paradigm shift in academic publishing was initially rejected wholesale by many in the research community but the landscape has now been transformed and open access publishing is standard practice. Open access is about the democratisation of information; the new approach to data sharing being taken forward by InterConnect extends this to the level of data, sharing data to create new knowledge and doing so in a democratic manner by which all can lead analysis rather than just contribute data for others to analyse. Parallels also exist in the actions that need to happen to create the paradigm shift: an initial political will and acceptance of a direction of travel; identification of barriers and obstacles to implementation and that funder / researcher behaviour is altered by infrastructure changes and alterations in incentives. The importance of the patient voice through this change process was also emphasised.

Professor Wareham drew the meeting to a close, and thanked participants for attending and for their welcome contributions to the meeting. It was agreed that participants would be contacted in the future as a sounding board, forming the nucleus of a virtual funders' and stakeholder network to begin a dialogue to develop a shared vision of InterConnect.

Photographs of the participants during the workshop.

SPEAKER BIOGRAPHIES

Dr Nick Banatvala, Senior Adviser to the Assistant Director General, Noncommunicable Diseases and Mental Health, World Health Organization, Geneva, Switzerland. Dr Banatvala is responsible for leading development of a global coordination mechanism for the prevention and control of NCDs, spearheading a newly set up UN NCD Taskforce and leading WHO's global training programme to build capacity on NCDs for senior policy makers in middle and low-income countries. Dr Banatvala trained in paediatrics and infectious diseases, then did public health and epidemiologic research in the UK and at CDC, Atlanta. He has sat on government, non-government and academic boards, as well as national and international committees and has undertaken consultancies for agencies including the World Bank.

Dr Karim Berkouk, Deputy Head of Unit of the European Commission Medical Research Unit in the Health Directorate of the Research and Innovation DG, Brussels, Belgium. Dr Berkouk studied Fluid Mechanics at the University of Paul Sabatier (Toulouse, FR) before completing his PhD at the University of Warwick (UK) in bio-fluid mechanics, where he worked on syringomyelia, a rare disease of the spinal cord In 2004, he joined the European Commission to work for the Marie Curie programme first as a project officer and then as a Head of Sector. In November 2011, he joined the Medical Research unit in the Health Directorate as the deputy head of unit to develop and implement research policies on cancer, brain, cardiovascular, chronic diseases, ageing, diabetes and obesity.

Dr Isabel Fortier, Director of the Maelstrom-Research Program at the Research Institute of McGill University Health Center, Montreal, Canada. Dr Fortier leads the Maelstrom Research and DataSHaPER programs at the Research Institute of the McGill University Health Centre. These initiatives aim to develop and provide the research community from diverse disciplines with resources (expertise, methods and software) to leverage and support data harmonization and integration across studies and implementation of individual study infrastructures. In addition, she serves as coordinator of data harmonization for the BioSHaRE (Biobank Standardisation and Harmonisation for Research Excellence in the European Union) project.

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 qualified in Medicine from Newcastle University and trained in Epidemiology at the London School of Hygiene and Tropical Medicine as a Wellcome Clinical Training Fellow, and in and Public Health Medicine in London and Cambridge. 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 Ruth Loos, Director, Genetics of Obesity and Related Metabolic Traits Program, The Ichan School of Medicine at Mount Sinai, New York, USA. Dr Loos obtained her PhD at the University of Leuven, Belgium, after which she was a postdoctoral fellow in Dr Claude Bouchard's Human Genetics laboratory at the Pennington Biomedical Research Center, Baton Rouge, USA. In 2005, she joined the MRC Epidemiology Unit of the Institute of Metabolic Science in Cambridge to become Group Leader of the Genetic Aetiology of Obesity Programme. Dr Loos joined the Mount Sinai School of Medicine in New York, and remains an honorary member of the MRC Epidemiology Unit in Cambridge. Her primary research interests focus on the identification of genes and genetic loci contributing to the risk of obesity and related metabolic traits.

Professor Matthias Schulze, Head, Department of Molecular Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke, Germany. Prof Schulze studied Nutritional Sciences at the Friedrich Schiller University, Jena, Germany and Public Health at Tulane University, New Orleans. He got his PhD training at the Technical University Berlin in cooperation with the German Institute of Human Nutrition (DIfE). Prof Schulze was appointed as full Professor at the University Potsdam and head of the department of Molecular Epidemiology at DIfE in 2010. His research interests are methodological approaches for dietary pattern analyses, metabolic and genetic predictors of type 2 diabetes, interaction between genetic factors and diet in the development of type 2 diabetes, and risk prediction modelling.

Professor Ronald Stolk, Head of Department and Chief Scientific Officer LifeLines, Department of Epidemiology, University Medical Center, Groningen, The Netherlands. Prof Stolk's research focuses on life course epidemiology approaches of chronic diseases, based on cohort studies and gene-environment interactions. Life Course Epidemiology is a central aspect of the Healthy Ageing theme of the University Medical Center Groningen. He did his medical training in Rotterdam and was trained in epidemiology at the universities of Rotterdam, Utrecht and Sydney. Prof Stolk has become involved in infrastructure for cohort studies. These biobanks include population based studies (LifeLines, GECKO, PIAMA, BioSHaRE) as well as clinical follow- up projects. He is also involved in harmonization of data from cohort studies, including cataloguing. Harmonization allows pooling of data from different studies, required to investigate complex interaction of multiple genetic and/or lifestyle factors.

Professor Nick Wareham, Director MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, UK. Prof Wareham is an Honorary Consultant at Addenbrooke's Hospital, Cambridge and Co-Director of the Institute of Metabolic Science. He was formerly a Wellcome Trust Senior Fellow in Clinical Science in Cambridge. He qualified in medicine from St Thomas' Hospital Medical School, London, and trained in epidemiology and public health at the London School of Hygiene and Tropical Medicine, London, Harvard University, Boston, USA and at the University of Cambridge. He was the coordinator of the InterAct EU FP6 funded project, which investigated how genes and lifestyle factors interact to lead to type 2 diabetes. He has a programme of research into the genetic basis of obesity and type 2 diabetes as part of the genetics of energy metabolism consortium. His work on gene-environment interaction is based on quantitative trait studies and large scale population-based cohort studies Prof Wareham is co-lead of the ADDITION study, a trial of screening for diabetes and intensive cardiovascular risk reduction undertaken in three European Countries.