InterConnect

Global data for diabetes and obesity research

Towards a global initiative on gene-environment interactions on diabetes and obesity

Nick Wareham
Director MRC Epidemiology Unit, University of Cambridge, UK
Brussels, 10th October 2014

This project is funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 602068.

Programme of the day

- Session 1 – Setting the scene
- Session 2 – Challenges of current data sharing models
- Session 3 – Vision of a changed paradigm
- Session 4 – Next steps – what can we do to move towards this changed paradigm

Session 1: Setting the Scene

Perspectives from the WHO on global epidemic of diabetes/obesity
Nick Banatvala

The EU “diabesity” conference 2012
Nick Wareham

GACD as a vehicle for intervention research
Karim Berkouk

Understanding differences in risk of diabetes and obesity between populations
Nick Wareham

Session 1: Setting the Scene

Perspectives from the WHO on global epidemic of diabetes/obesity
Nick Banatvala

The EU “diabesity” conference 2012
Nick Wareham

GACD as a vehicle for intervention research
Karim Berkouk

Understanding differences in risk of diabetes and obesity between populations
Nick Wareham

InterConnect meeting: perspectives from the World Health Organization

Dr Nick Banatvala
Senior Adviser to the Assistant Director General,
Noncommunicable Diseases and Mental Health
World Health Organization, Geneva

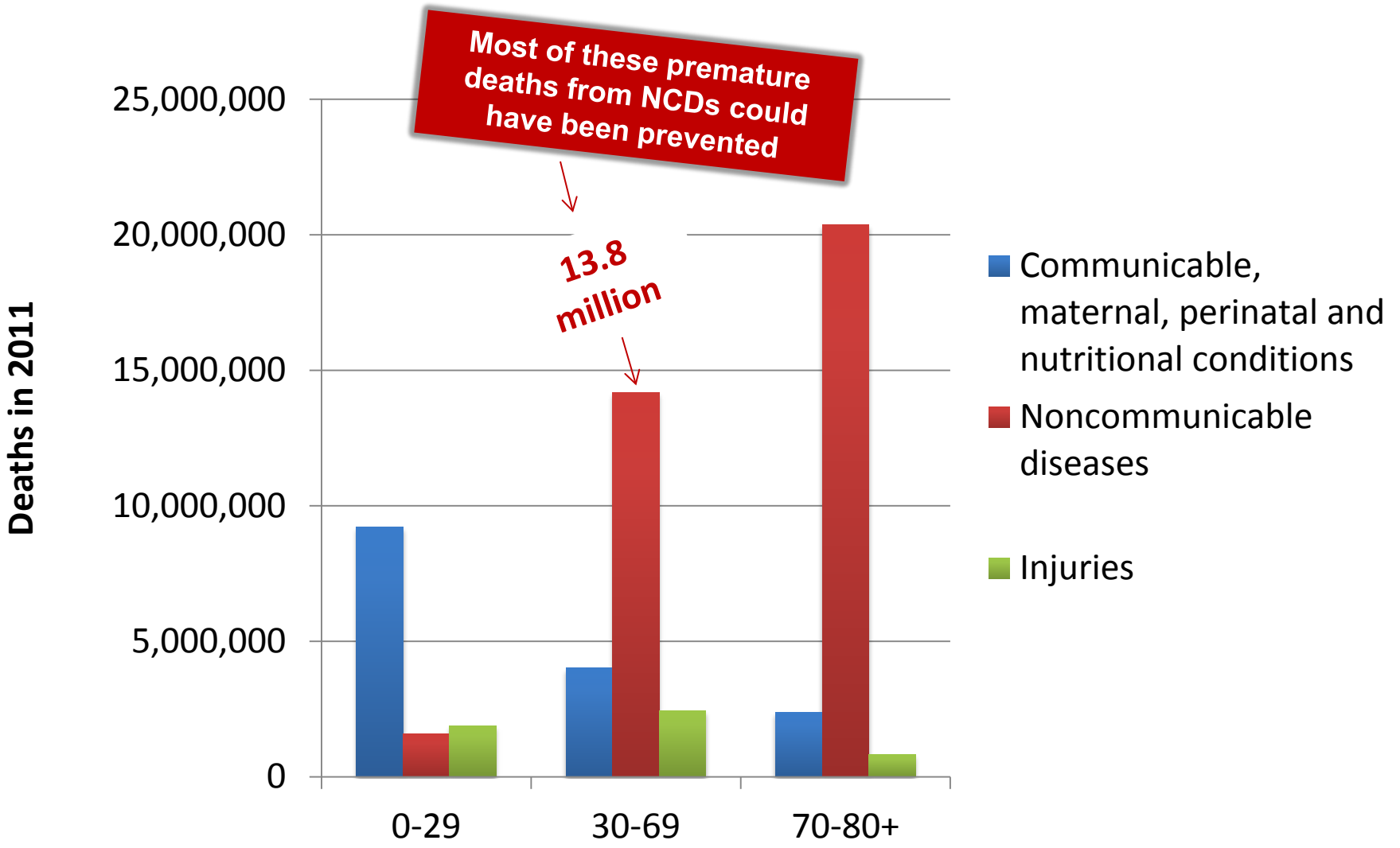
World Health Organization

UNITE IN THE FIGHT AGAINST NCDs

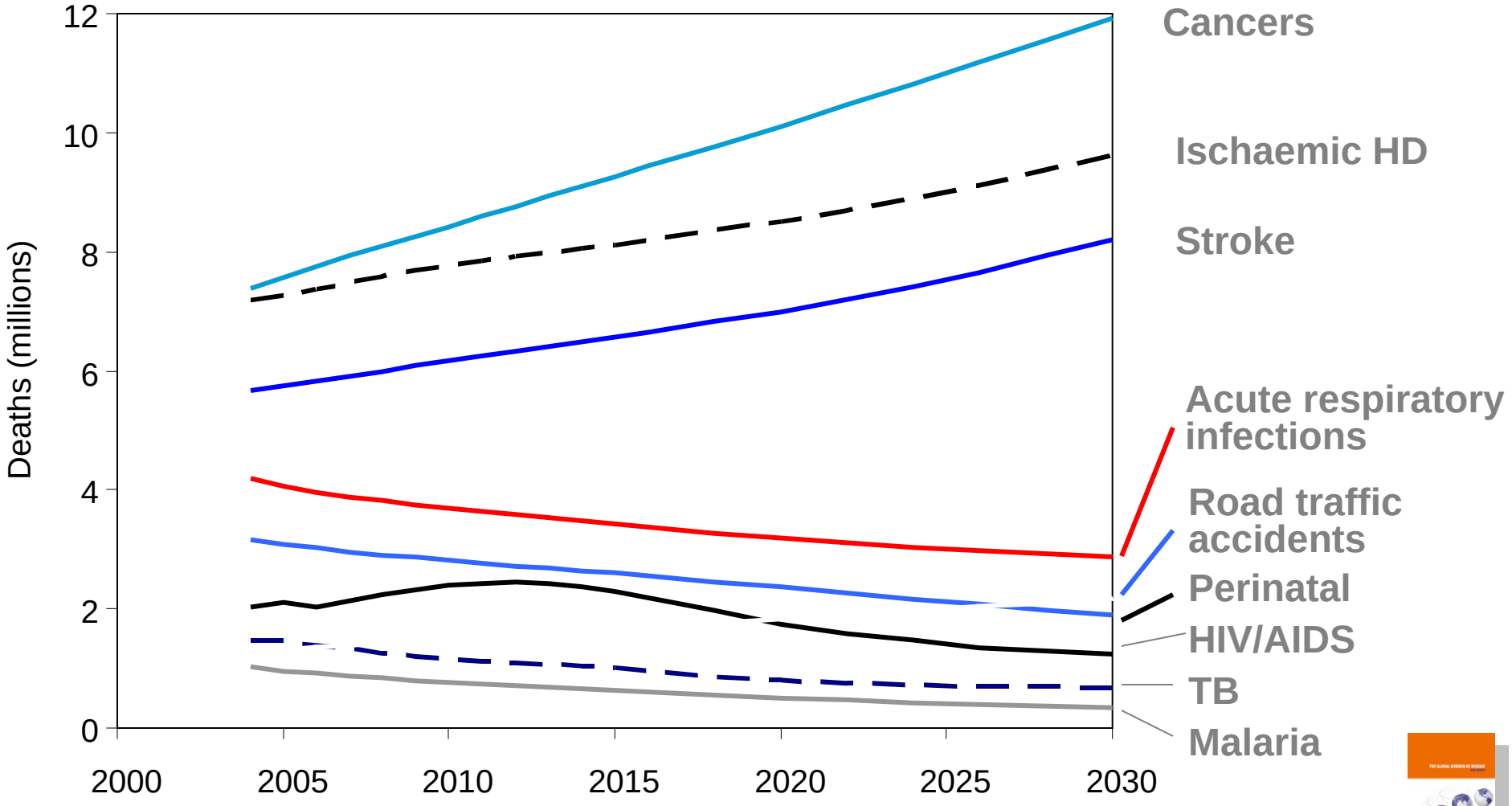
❶ ② ③ ④

NCDs represent one of the world's major development challenges, both in terms of the great human suffering they cause in all countries, as well as the immense harm they inflict on the socio-economic fabric of many countries, particularly the world's poorest

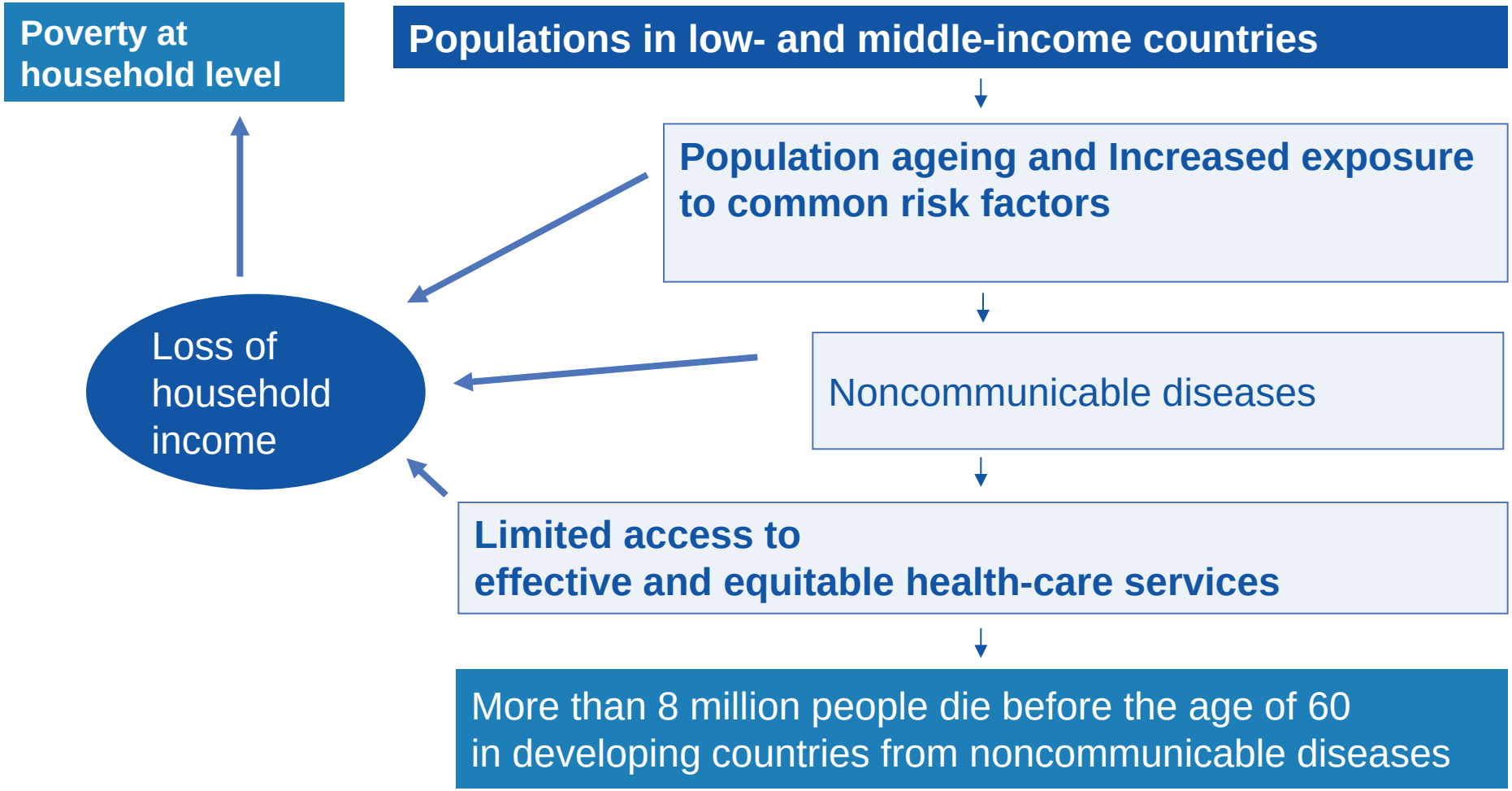
In 2011, 13.8 million people died around the world from NCDs between the ages of 30 and 70: more than 85% of these deaths occurred in developing countries



Global projections (2004 to 2030)

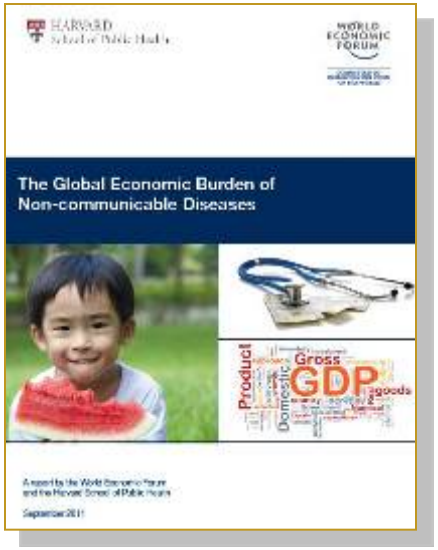


NCDs and Development
Poverty contributes to NCDs and NCDs cause poverty



Economics

The cost of inaction versus action and the costs of scaling up



US$ 7T

cumulative lost output in developing countries associated with NCDs between 2011-2025

US$ 11B

average yearly cost for all LMICs to scale up action by implementing the "best buys"

US$1 per capita in LICs

US$1.5 and US$3 in LMICs and UMICs

① ② ③ ④

Diabetes is a major contribution to the burden of NCDs

Number of persons with diabetes in the world

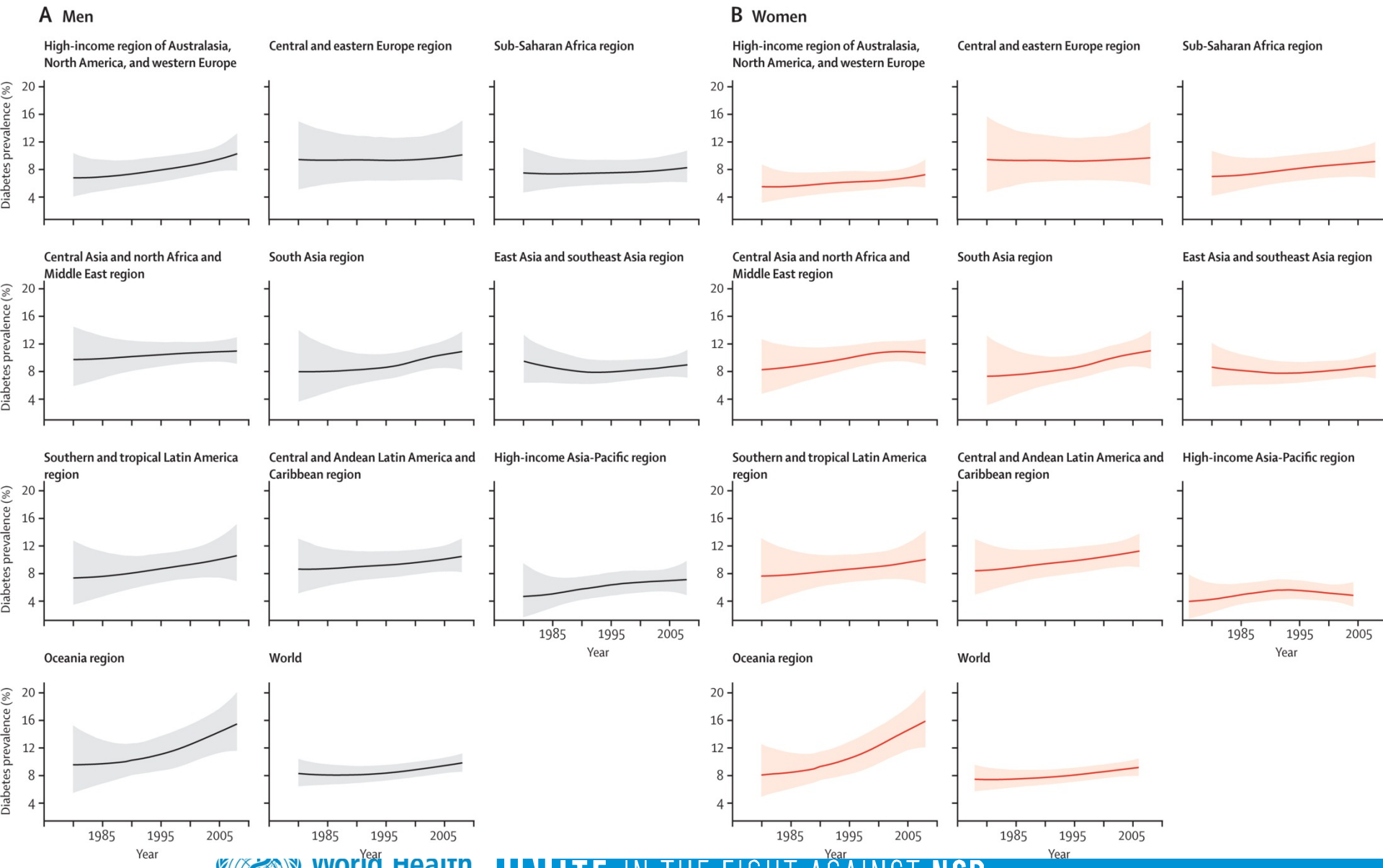
346 million in 2008 (WHO, 2011)

382 million in 2013 (IDF, 2013)

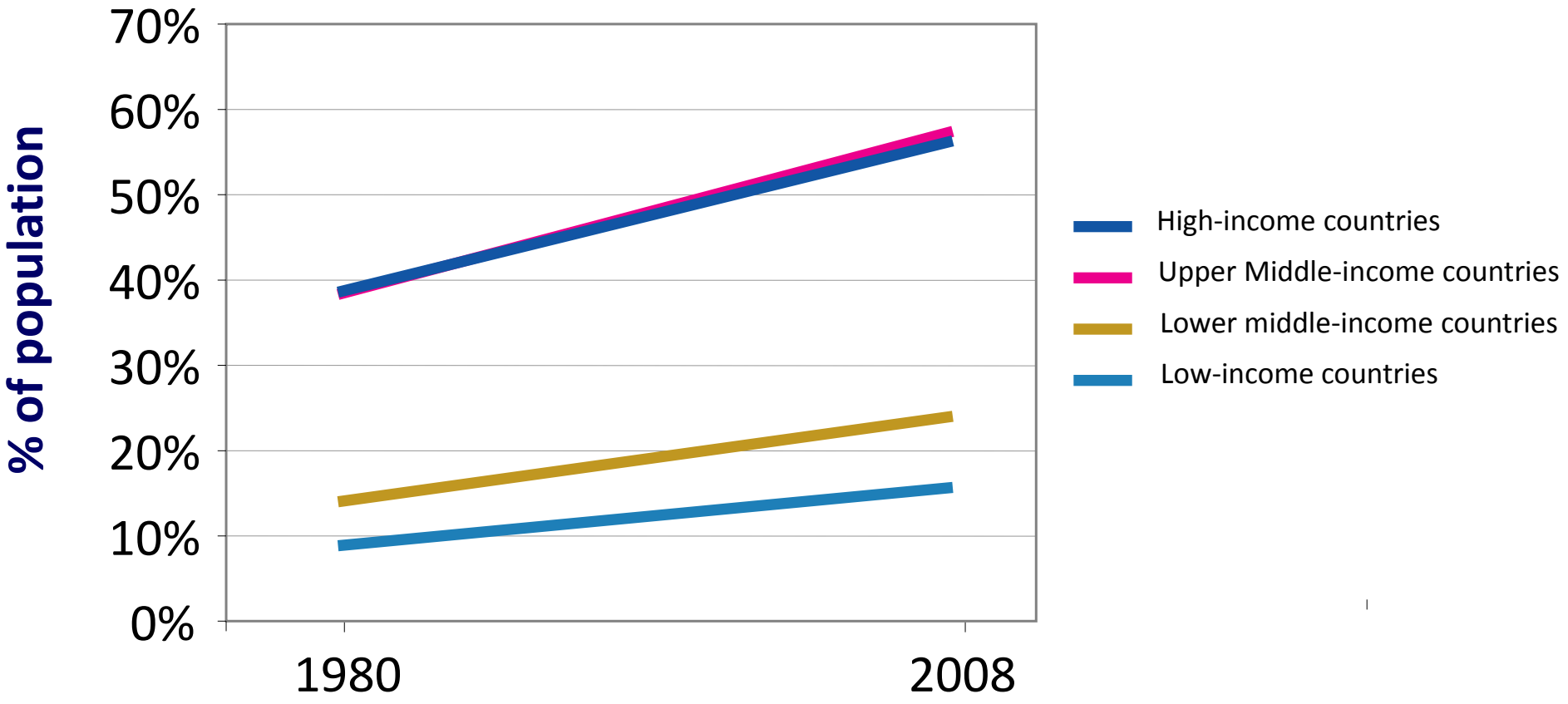
and in 2035…

592 million in 2035 (IDF, 2013)

Diabetes prevalence trend 1980-2008 (gbd lancet Danaei et al, 2011)

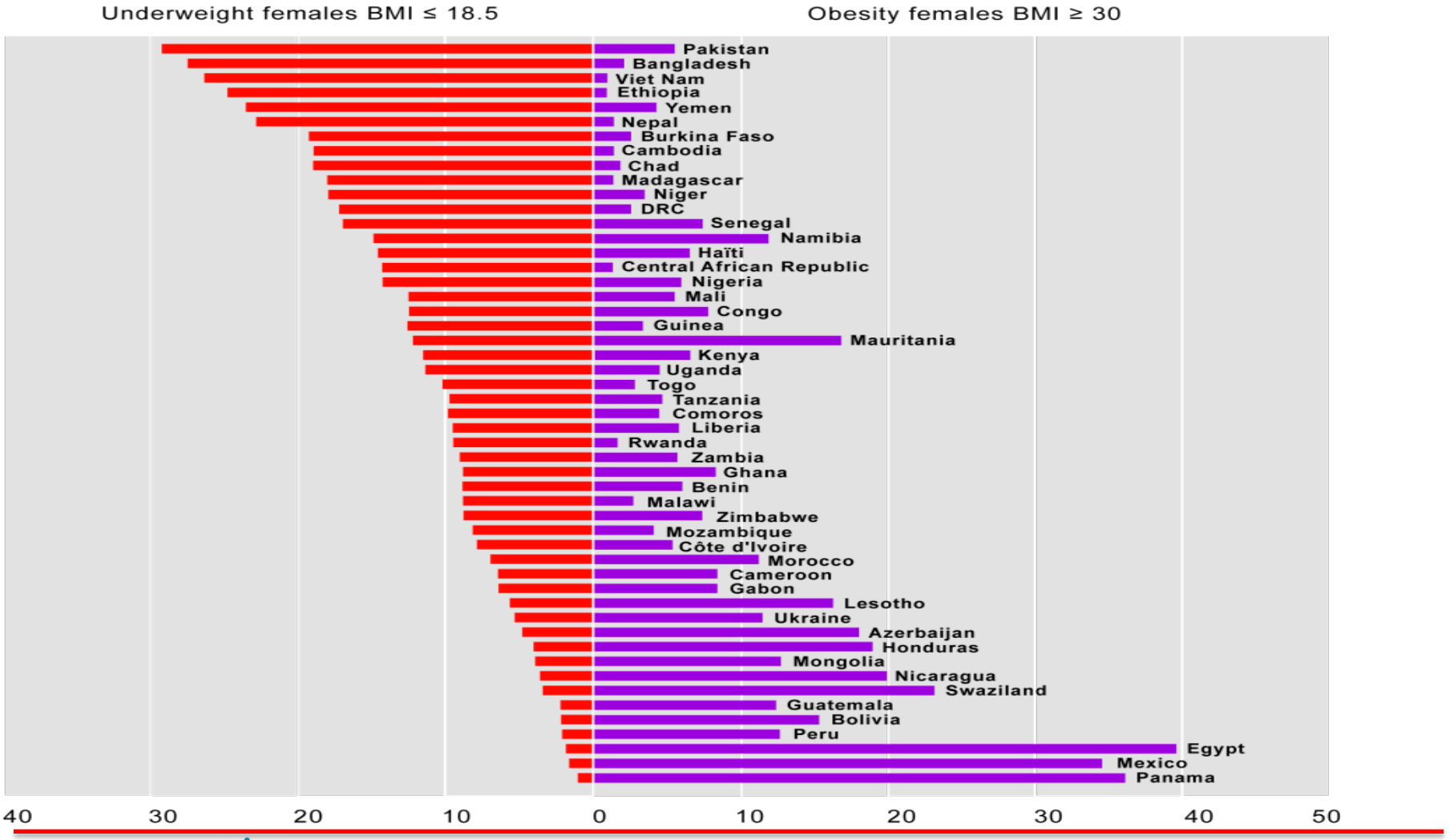


Overweight over the last 30 years



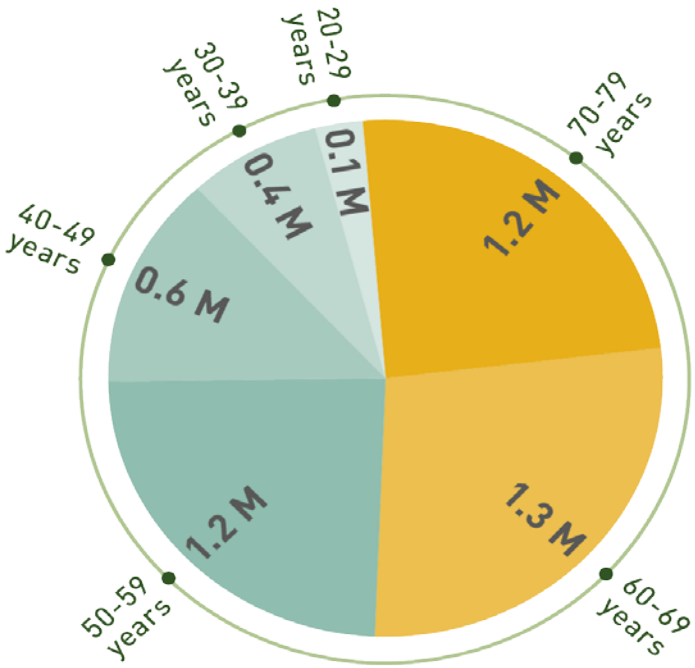
Overweight (2008)
Source: WHO NCD Country Profiles (2010)

Underweight and obesity coexisting in low and middle income countries



Half of people who die from diabetes are **under the age of 60.**

DEATHS ATTRIBUTABLE TO DIABETES BY AGE (20-79 YEARS)



IDF, 2013

The poorest people in developing countries affected the most

The cost of caring for a family member with diabetes can be more than 20 per cent of low-income household incomes in developing countries

The cost per year of diabetes care at household level

	Insulin	Syringes	Testing	Consultation	Travel	Total cost	% of per capita Income
Mali (2004)	38%	34%	8%	7%	12%	$339.4	61%
Mozambique (2003)	5%	24%	1%	9%	61%	$273.6	75%
Nicaragua (2007)	0%	73%	0%	0%	27%	$74.4	7%
Zambia (2003)	12%	63%	6%	6%	12%	$199.1	21%
Vietnam (2008)	39%	8%	5%	3%	46%	$427.0	51%

1 2 3 ④

Responding to the challenge of diabetes

The United Nations is addressing NCDs as one of the major challenges for development

2000 — Global Strategy for the Prevention and Control of NCDs

2003 — FCTC WHO FRAMEWORK CONVENTION ON TOBACCO CONTROL

Global Strategy on Diet, Physical Activity and Health

2004

2008-2013 Action Plan on the Global Strategy for the Prevention and Control of NCDs

2008

Global Strategy to Reduce the Harmful Use of Alcohol

2009

First WHO Global Status Report on NCDs

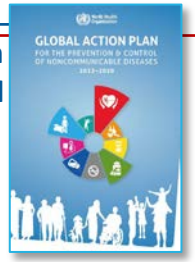
2010

Moscow Declaration

2011 — 2011 UN Political Declaration on NCDs

UN Task Force on NCDs

2013 — WHO Global NCD Action Plan 2013-2020, including 9 global targets and 25 indicators



2014 — 2014 UN Outcome Document on NCDs

2015 — Country Framework for Action to engage sectors beyond health on NCDs

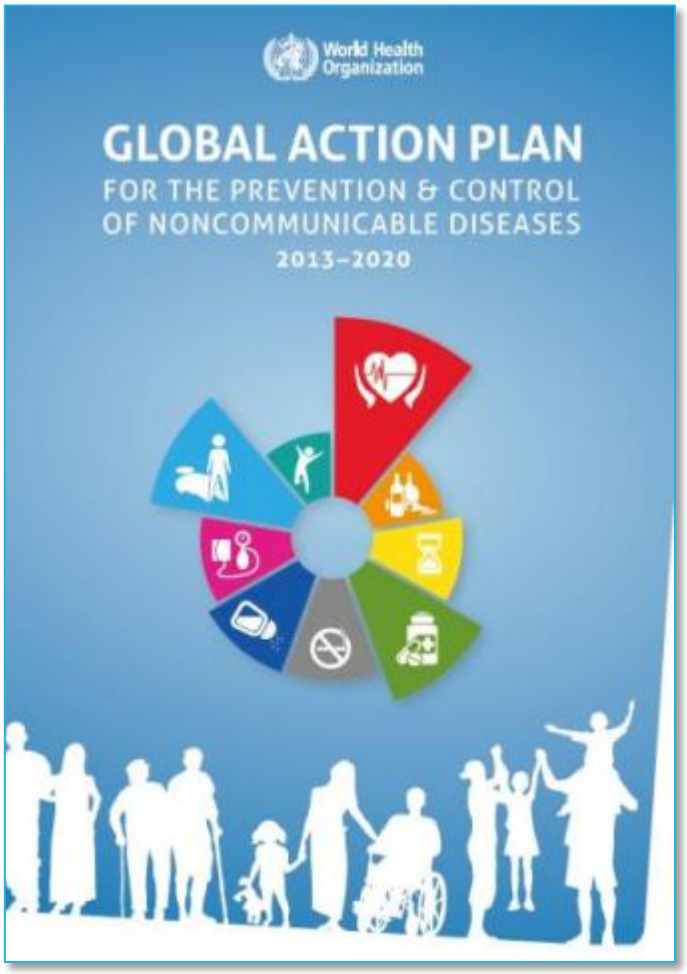
Adoption of the Post-2015 development agenda

2025 — Attainment of the 9 global targets for NCDs by 2025

2030 — Attainment of the 9 global targets for NCDs by 2030 (as part of the post-2015 development agenda)

- World Health Assembly
- UN General Assembly
- ECOSOC

WHO Global NCD Action Plan 2013-2020



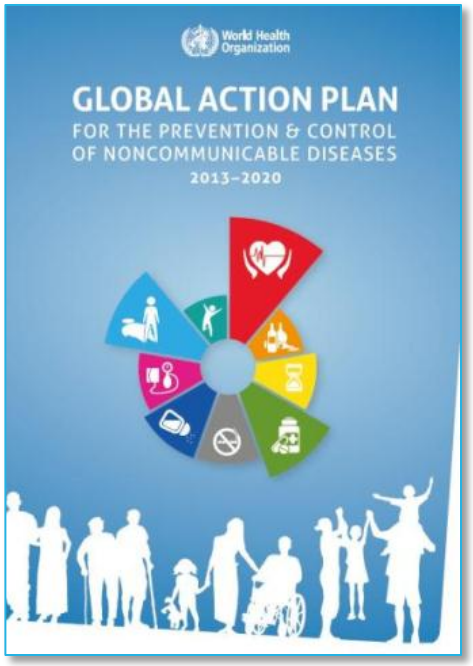
Vision:

A world free of the avoidable burden of NCDs

Goal:

To reduce the preventable and avoidable burden of morbidity, mortality and disability due to NCDs by means of multisectoral collaboration and cooperation at national, regional and global levels

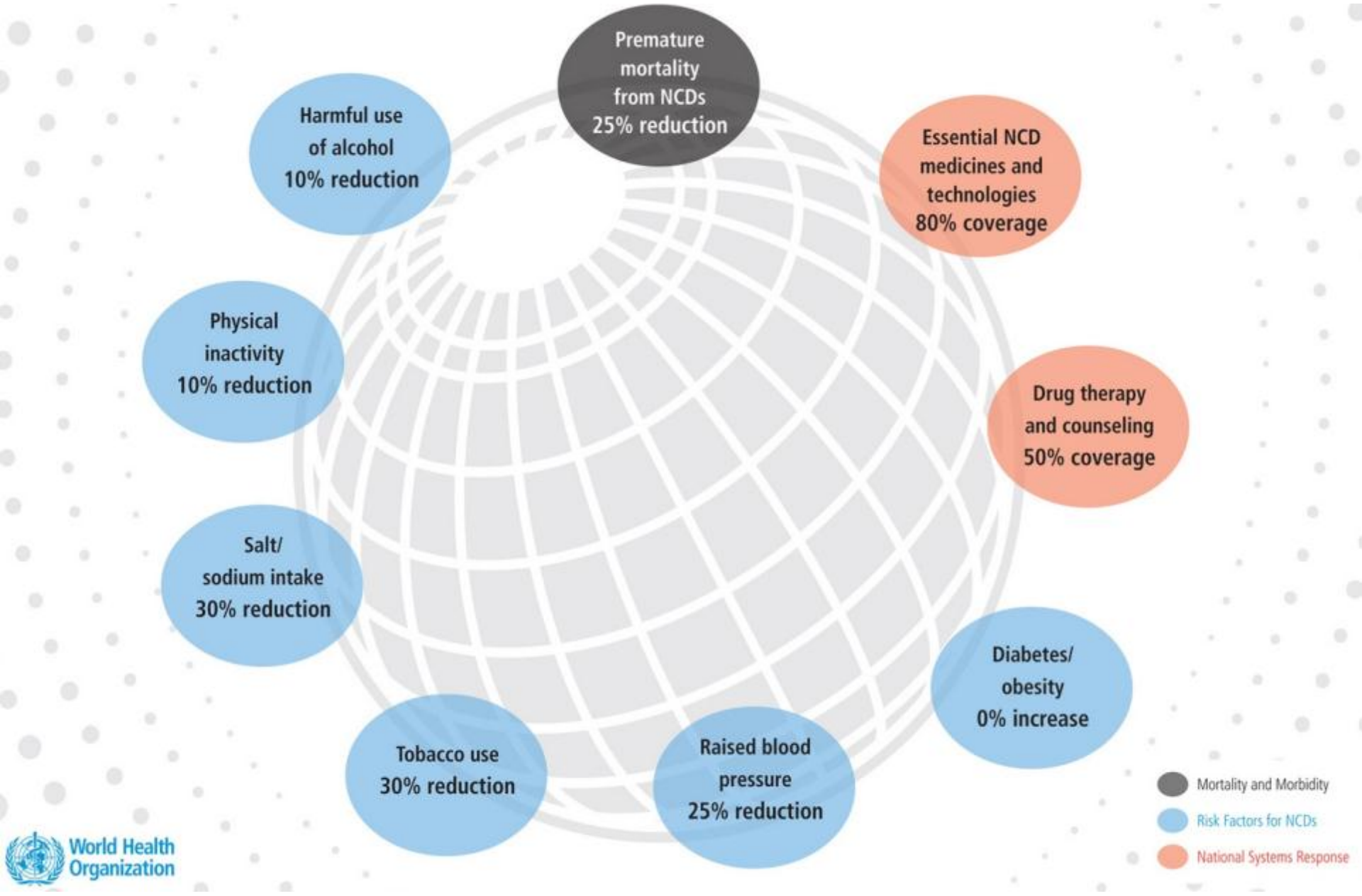
WHO Global NCD Action Plan 2013-2020



Objectives:

1. Raising the priority
2. National capacity, leadership and multisectoral action
3. Modifiable risk factors
4. Health systems
5. National capacity for high-quality research
6. Monitoring trends &determinants of NCDs

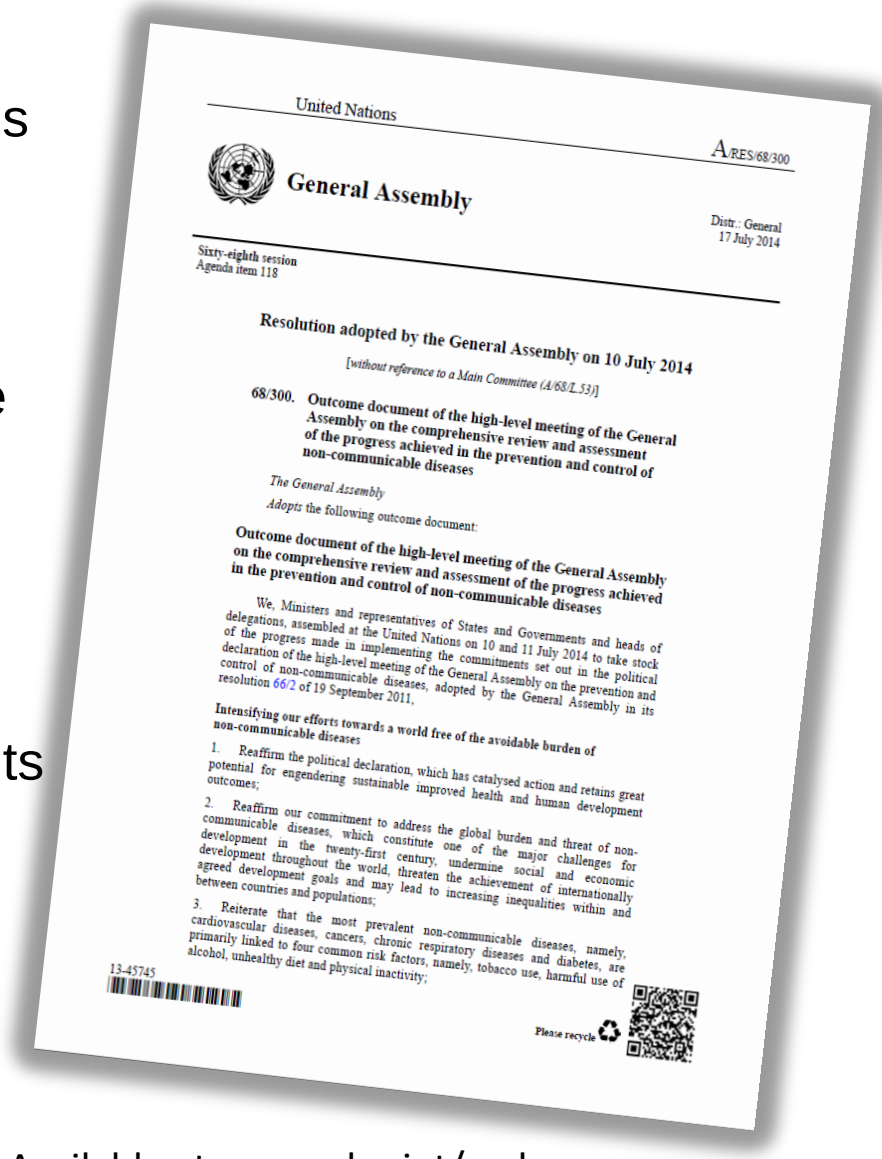
9 global targets to be attained by 2025



World Health Organization

2014 UN Outcome Document on NDCs (resolution A/RES/68/300)

- **Bottom line:**
  Governments are committed themselves to intensify their efforts towards a world free of the avoidable burden of NCDs
- **Taking stock:**
  Acknowledges progress achieved since 2011
- **Reaffirming our leadership:**
  Reiterates existing commitments
- **Moving forward:**
  Maps out concrete national commitments between 2014 and 2018
- **Moving forward:**
  Provides 3 new global assignments
- **Towards the world we want:**
  Next milestone in 2018



United Nations A/RES/68/300

General Assembly

Distr.: General
17 July 2014

Sixty-eighth session
Agenda item 118

Resolution adopted by the General Assembly on 10 July 2014

[without reference to a Main Committee (A/68/L.53)]

68/300. Outcome document of the high-level meeting of the General Assembly on the comprehensive review and assessment of the progress achieved in the prevention and control of non-communicable diseases

The General Assembly

Adopts the following outcome document:

Outcome document of the high-level meeting of the General Assembly on the comprehensive review and assessment of the progress achieved in the prevention and control of non-communicable diseases

We, Ministers and representatives of States and Governments and heads of delegations, assembled at the United Nations on 10 and 11 July 2014 to take stock of the progress made in implementing the commitments set out in the political declaration of the high-level meeting of the General Assembly on the prevention and control of non-communicable diseases, adopted by the General Assembly in its resolution 66/2 of 19 September 2011,

Intensifying our efforts towards a world free of the avoidable burden of non-communicable diseases

1. Reaffirm the political declaration, which has catalysed action and retains great potential for engendering sustainable improved health and human development outcomes;

2. Reaffirm our commitment to address the global burden and threat of non-communicable diseases, which constitute one of the major challenges for development in the twenty-first century, undermine social and economic development throughout the world, threaten the achievement of internationally agreed development goals and may lead to increasing inequalities within and between countries and populations;

3. Reiterate that the most prevalent non-communicable diseases, namely, cardiovascular diseases, cancers, chronic respiratory diseases and diabetes, are primarily linked to four common risk factors, namely, tobacco use, harmful use of alcohol, unhealthy diet and physical inactivity;

13-45745

Please recycle

Available at www.who.int/ncd

World Health Organization

UNITE IN THE FIGHT AGAINST NCDs

Prioritised research agenda for NCDs

- **Diabetes**

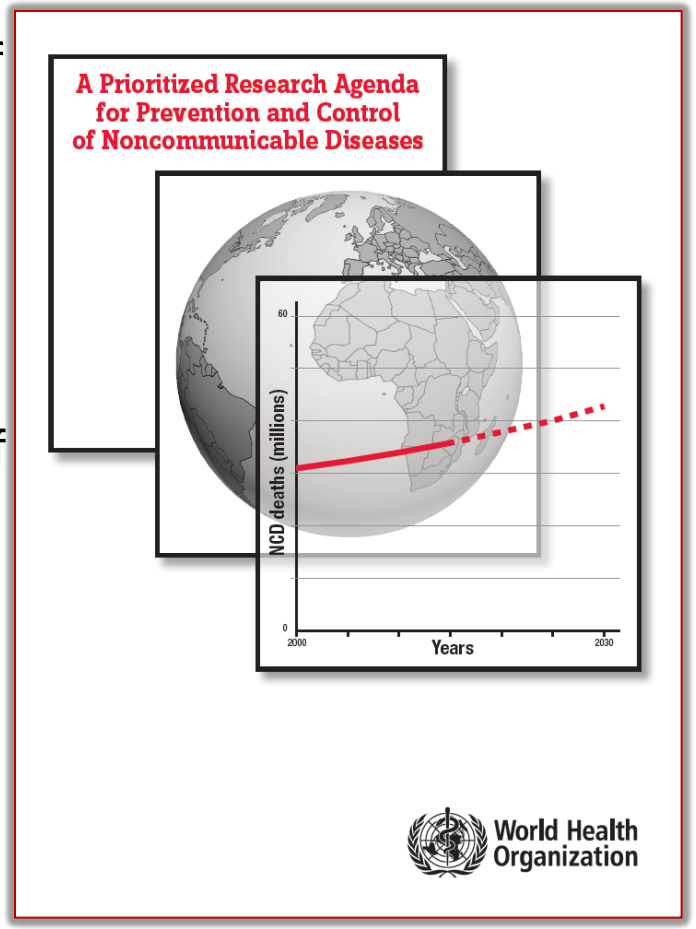
Three areas, one of which is identification of causes and measurement of magnitude

- **Nutrition and obesity**

Three areas, again one of which is identification of causes and measurement of magnitude

- **Genetics**

Three areas, one of which is analysis of problems and development of solutions



Available at www.who.int/ncd

1 2 3 4

InterConnect: an opportunity to explain better the problem and work more effectively

Session 1: Setting the Scene

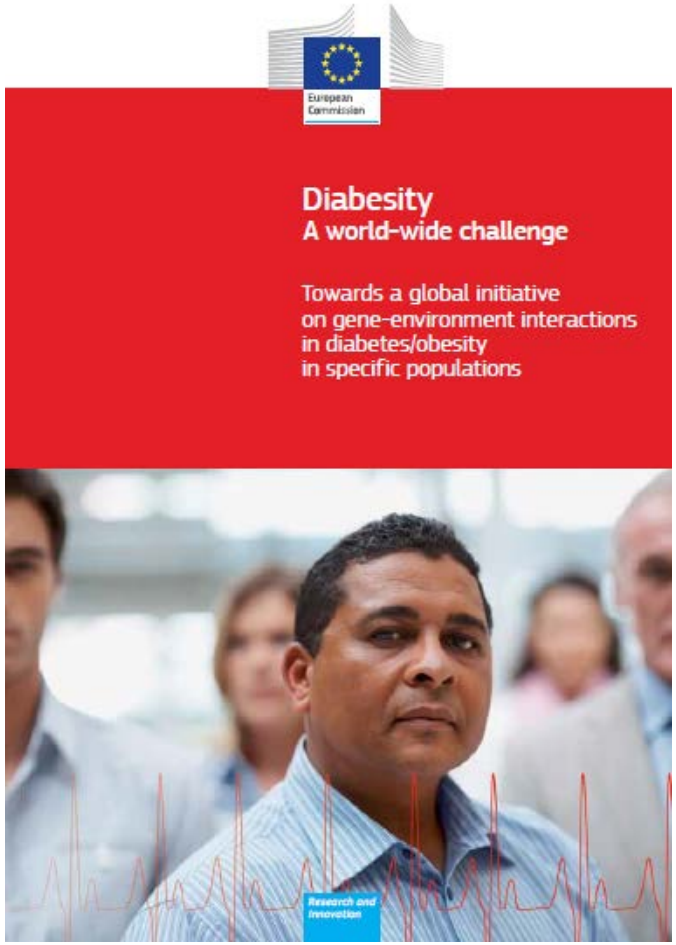
Perspectives from the WHO on global epidemic of diabetes/obesity
Nick Banatvala

The EU "diabesity" conference 2012
Nick Wareham

GACD as a vehicle for intervention research
Karim Berkouk

Understanding differences in risk of diabetes and obesity between populations
Nick Wareham

EU “diabesity” conference 2012



- Research into individual and societal approaches to the prevention of obesity, diabetes and related metabolic disorders
- Health systems interventions to better treat diabetes
- Research into understanding differences in individual and population risk

Session 1: Setting the Scene

Perspectives from the WHO on global epidemic of diabetes/obesity
Nick Banatvala

The EU “diabesity” conference 2012
Nick Wareham

GACD as a vehicle for intervention research
Karim Berkouk

Understanding differences in risk of diabetes and obesity between populations
Nick Wareham

Session 1: Setting the Scene

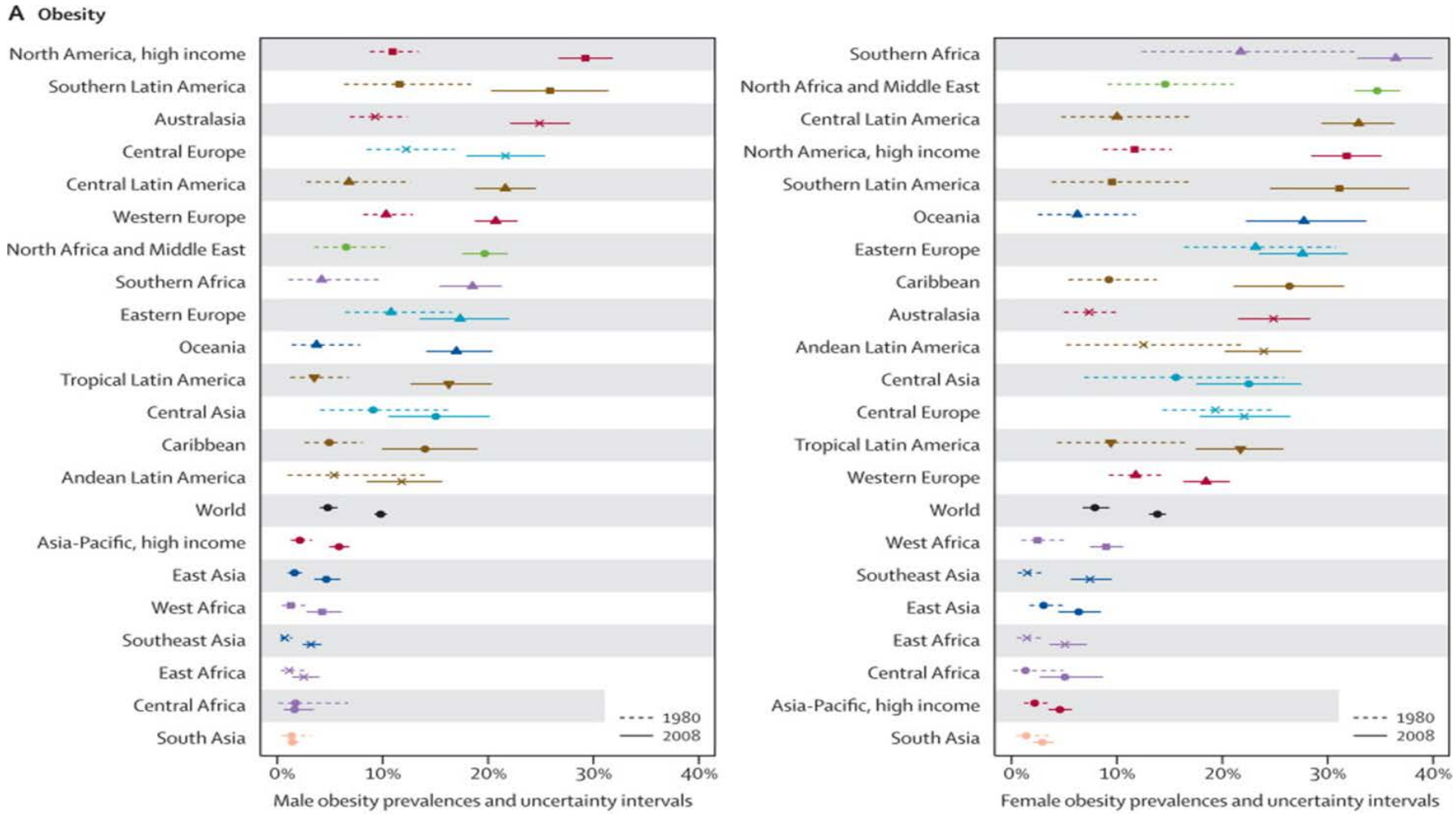
Perspectives from the WHO on global epidemic of diabetes/obesity
Nick Banatvala

The EU “diabesity” conference 2012
Nick Wareham

GACD as a vehicle for intervention research
Karim Berkouk

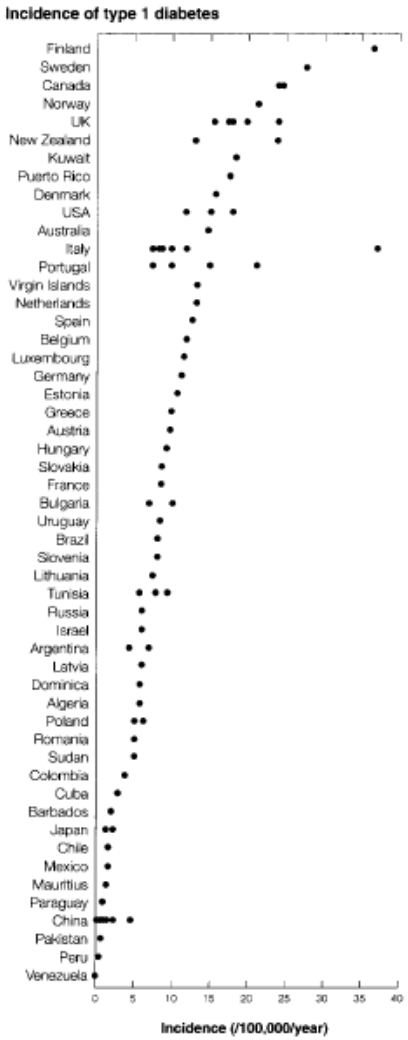
Understanding differences in risk of diabetes and obesity between populations
Nick Wareham

Phase 1: Describing difference in prevalence and incidence between populations



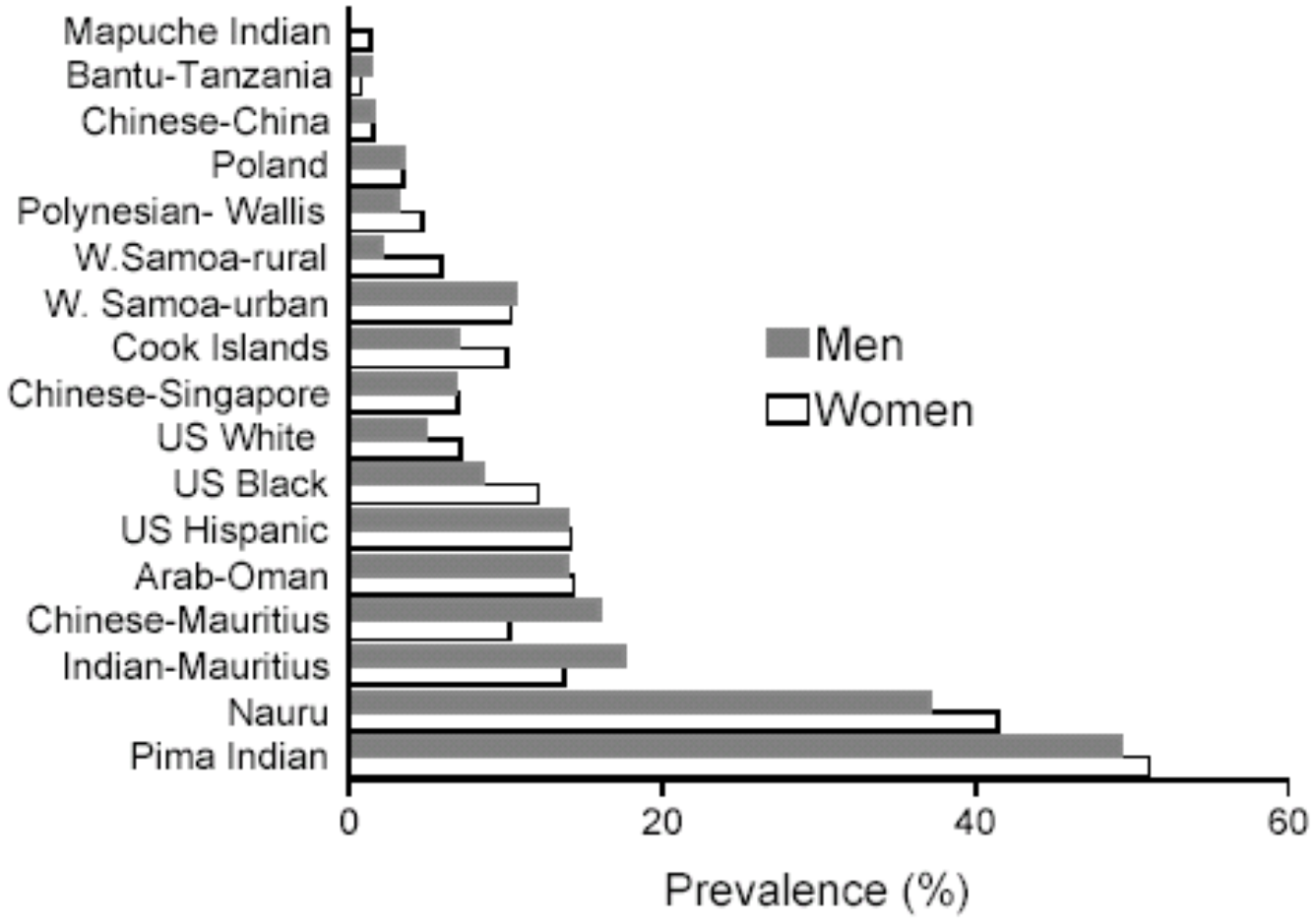
Source: Finucane et al, Lancet 2011

Between-population differences in incidence of type 1 diabetes



- High incidence in Finland, Sardinia and other populations
- On-going cohort studies in specific populations investigating interplay between genetic susceptibility and environmental triggers

Between-population differences in type 2 diabetes prevalence



Possible explanations for between-population differences in prevalence

THE AMERICAN JOURNAL
of
HUMAN
GENETICS

Diabetes Mellitus: A "Thrifty" Genotype Rendered Detrimental by "Progress"?

JAMES V. NEEL
Department of Human Genetics,
University of Michigan Medical School,
Ann Arbor, Mich.

Source: Neel, Am J Human Genetics 1962

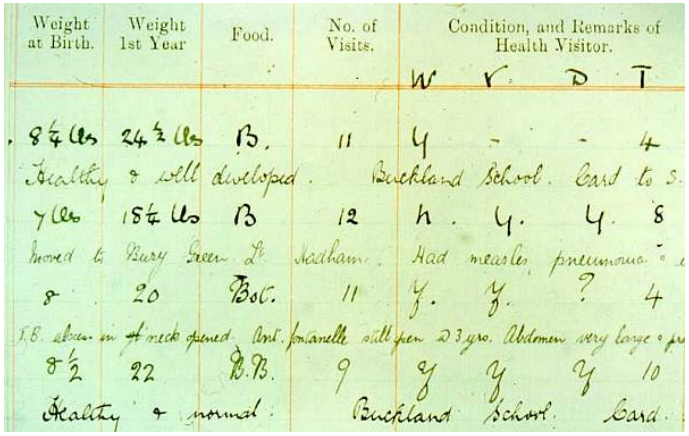
Possible explanations for between-population differences in prevalence

Review

Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis*

C.N. Hales[1] and D.J.P. Barker[2]

[1] Department of Clinical Biochemistry, Addenbrooke's Hospital, Cambridge, and
[2] MRC Environmental Epidemiology Unit, University of Southampton, Southampton General Hospital, UK



Source: Hales and Barker, Diabetologia 1992

Phase 2: Studying explanations for differences in risk between individuals within-populations

- EPIC-InterAct
  Nested case-cohort study within EPIC Europe

- Large
  455,680 individuals at baseline

- Long follow-up
  - 4 million person years
  - 12,403 incident cases of T2DM

- Stored blood
- Data on diet/physical activity
- Exposure heterogeneity

Design and cohort description of the InterAct Project: an examination of the interaction of genetic and lifestyle factors on the incidence of type 2 diabetes in the EPIC Study

The InterAct Consortium

Research groups in 8 countries; 26 centres

Source: Langenberg C et al, Diabetologia 2011

InterAct findings – foods associated with increased risk of T2DM

Diabetologia (2013) 56:47–59
DOI 10.1007/s00125-012-2718-7

ARTICLE

Association between dietary meat consumption and incident type 2 diabetes: the EPIC-InterAct study

The InterAct Consortium

Diabetologia (2013) 56:1520–1530
DOI 10.1007/s00125-013-2899-8

ARTICLE

Consumption of sweet beverages and type 2 diabetes incidence in European adults: results from EPIC-InterAct

The InterAct consortium

InterAct findings – foods associated with reduced risk of T2DM

The amount and type of dairy product intake and incident type 2 diabetes: results from the EPIC-InterAct Study[1–3]

Am J Clin Nutr 2012

The prospective association between total and type of fish intake and type 2 diabetes in 8 European countries: EPIC-InterAct Study[1–3]

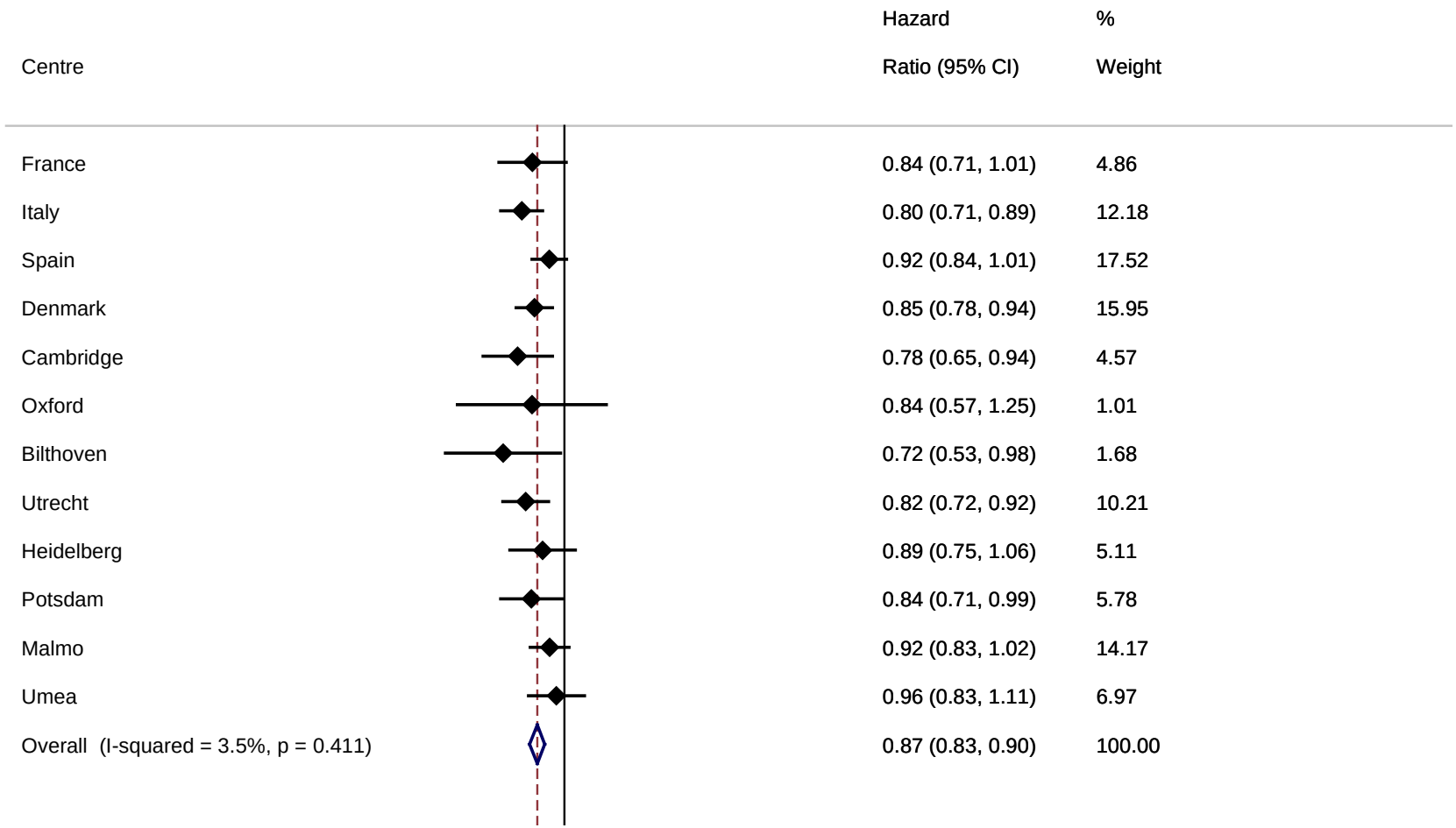
Am J Clin Nutr 2012

SYSTEMATIC REVIEW

Fruit and vegetable intake and type 2 diabetes: EPIC-InterAct prospective study and meta-analysis

European Journal of Clinical Nutrition (2012)

InterAct findings - Physical activity and risk of T2DM



Source: Ekelund et al, Diabetologia 2012

InterAct findings: Main genetic effect of known variants

49 variants previously demonstrated to be associated with T2DM

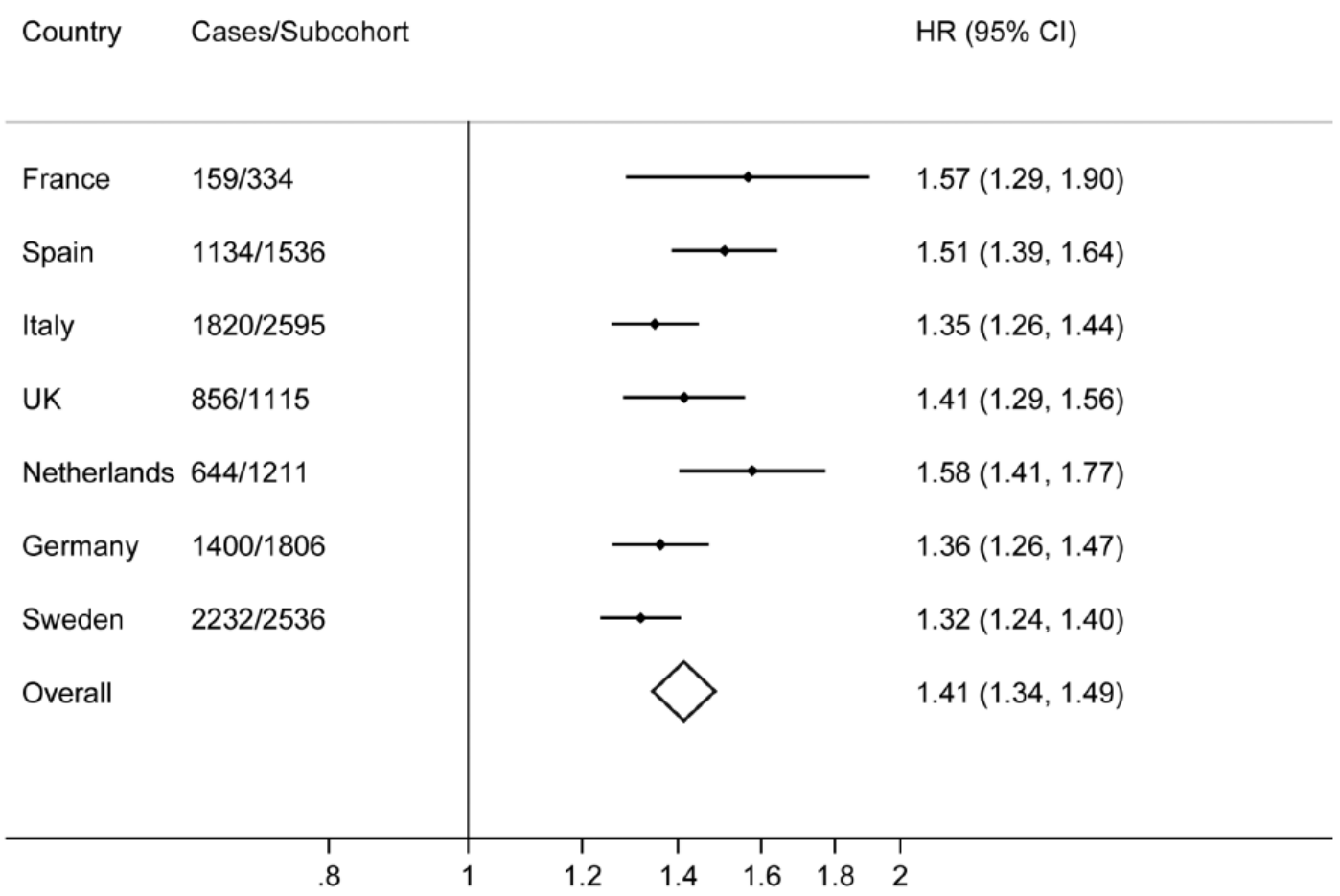
Genetic risk score strongly associated with incident T2DM – HR per allele 1.08 (1.07-1.10) $p = 10^{-41}$

Per SD of GRS HR = 1.41 (1.34-1.49) $p = 10^{-41}$

No evidence of interaction for individual gene variants with age, sex, family history, BMI or physical activity

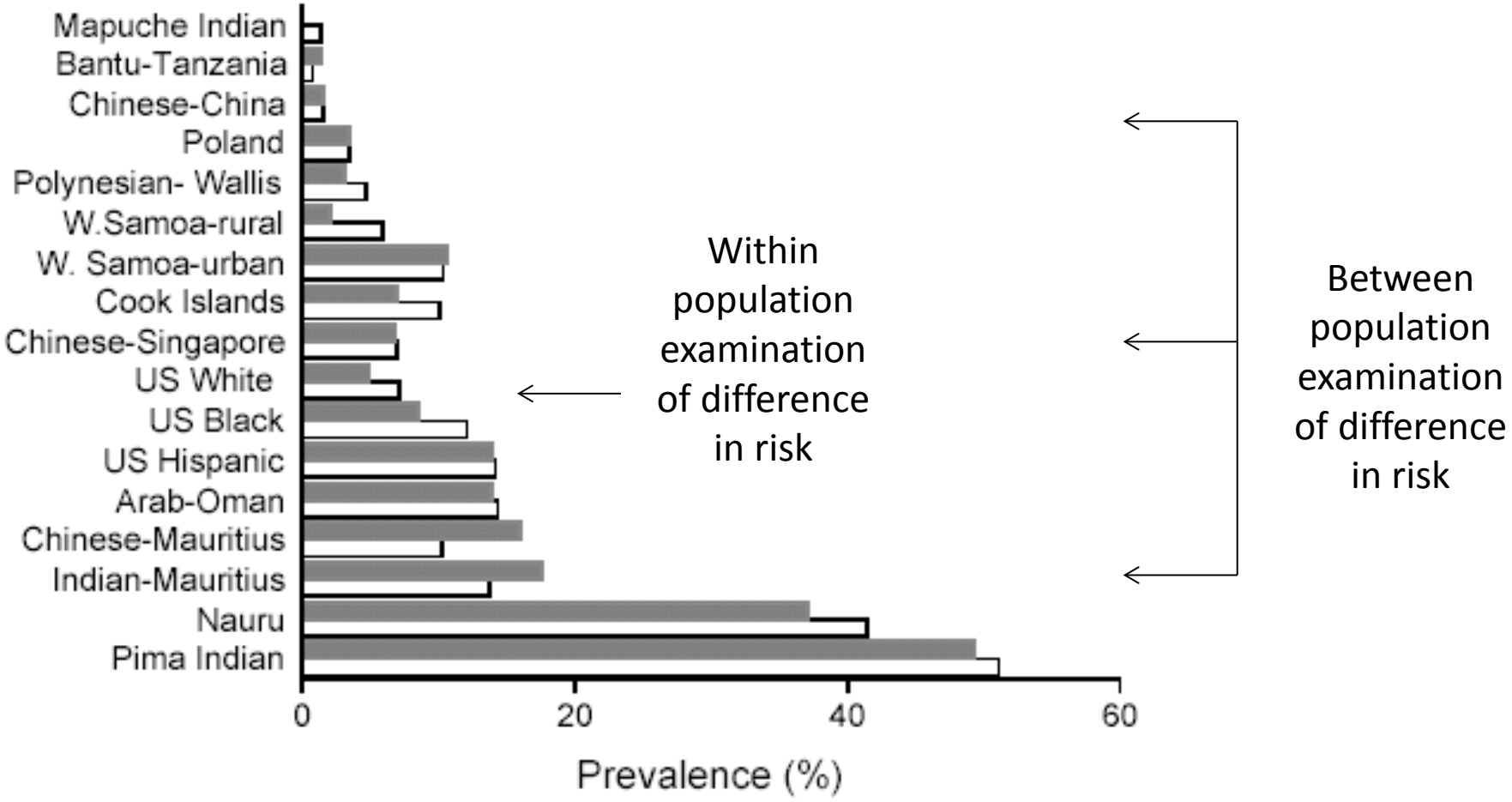
Source: Langenberg et al, PLoS Med 2014

InterAct findings: Main genetic effect by country



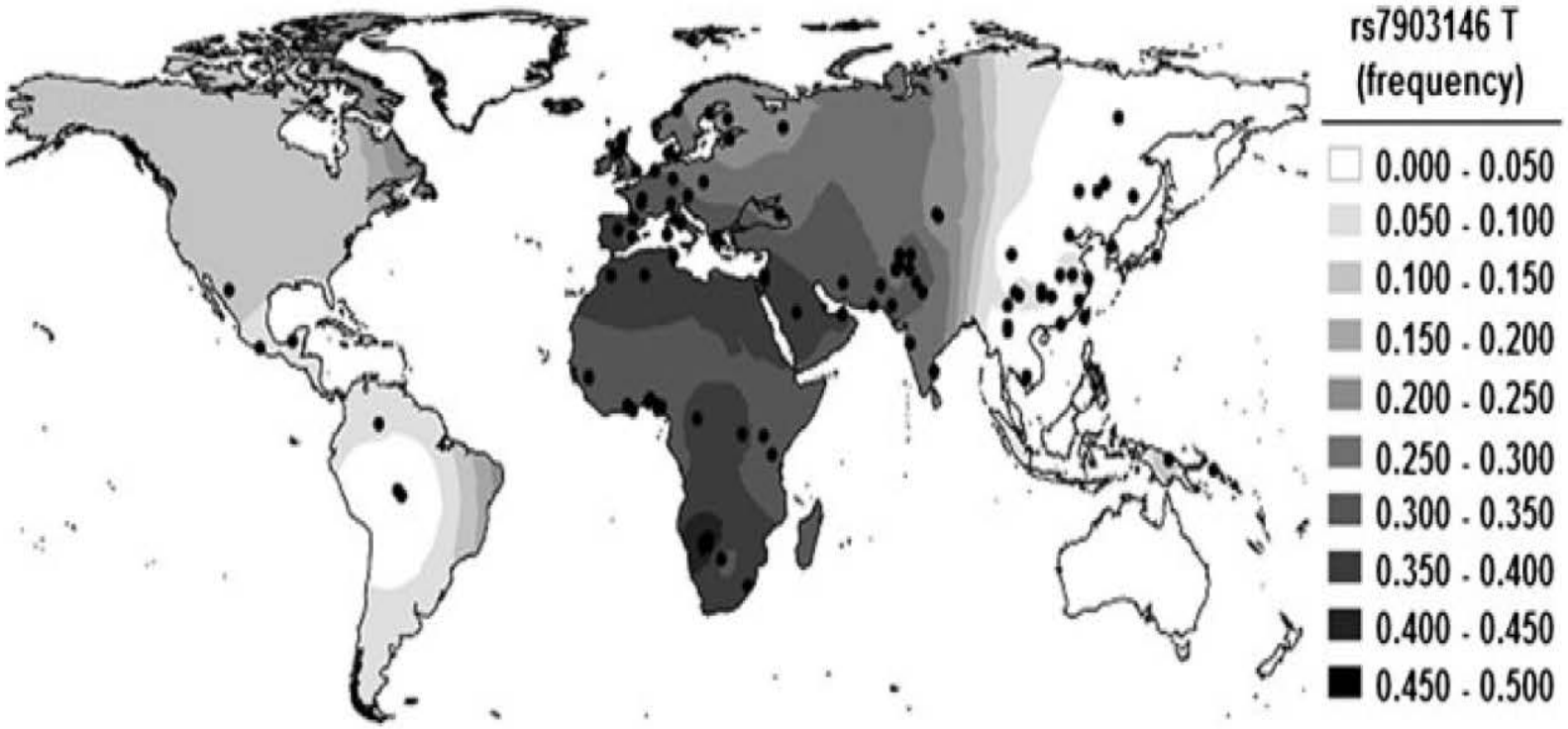
Source: Langenberg et al, PLoS Med 2014

Phase 3: Moving from within-population investigation to the study of between-population differences



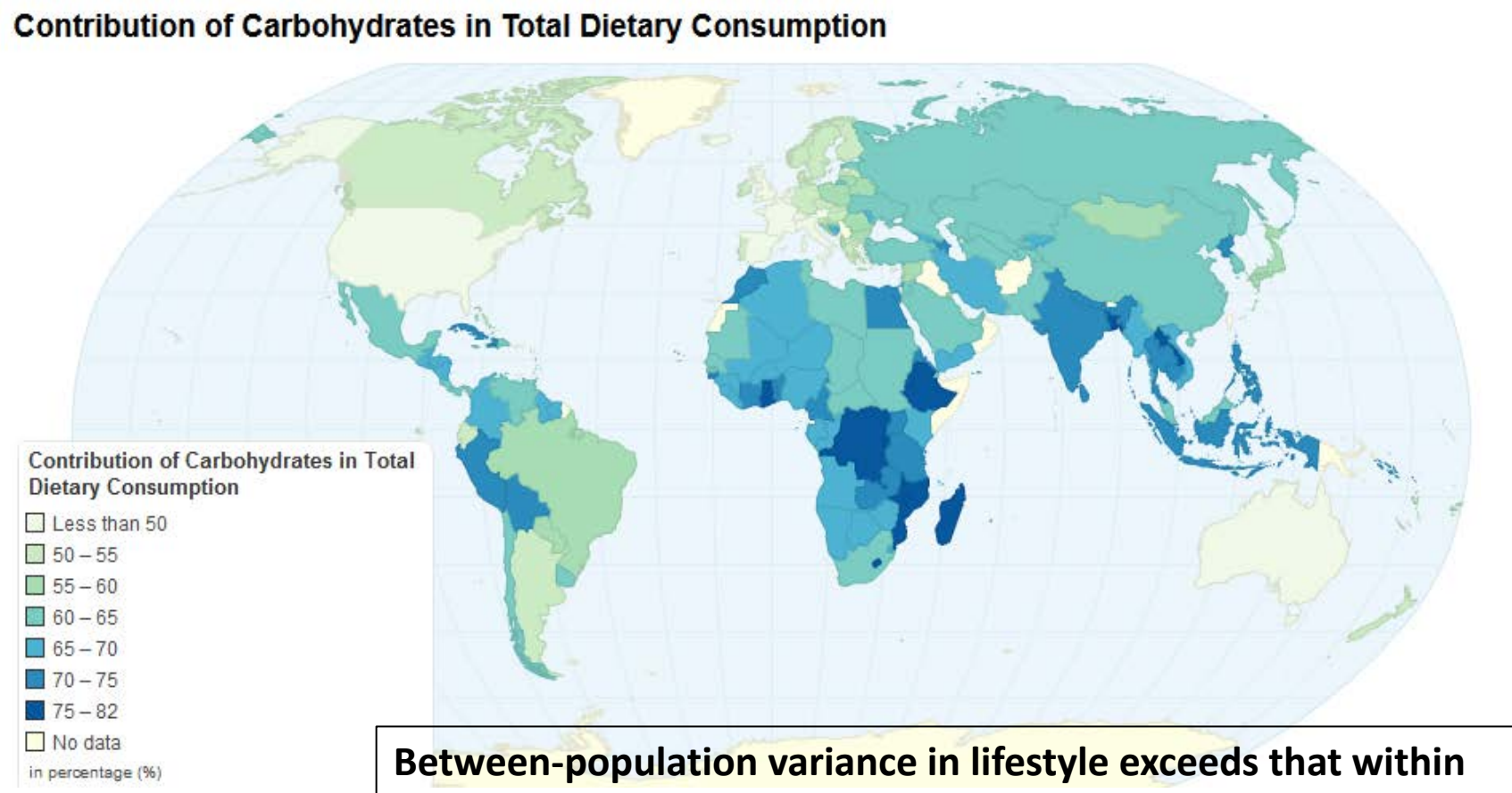
Studying between-population differences – genetics

Global distribution of rs7903146 T allele in TCF7L2



Source: Guinan, Biochem Genet 2012

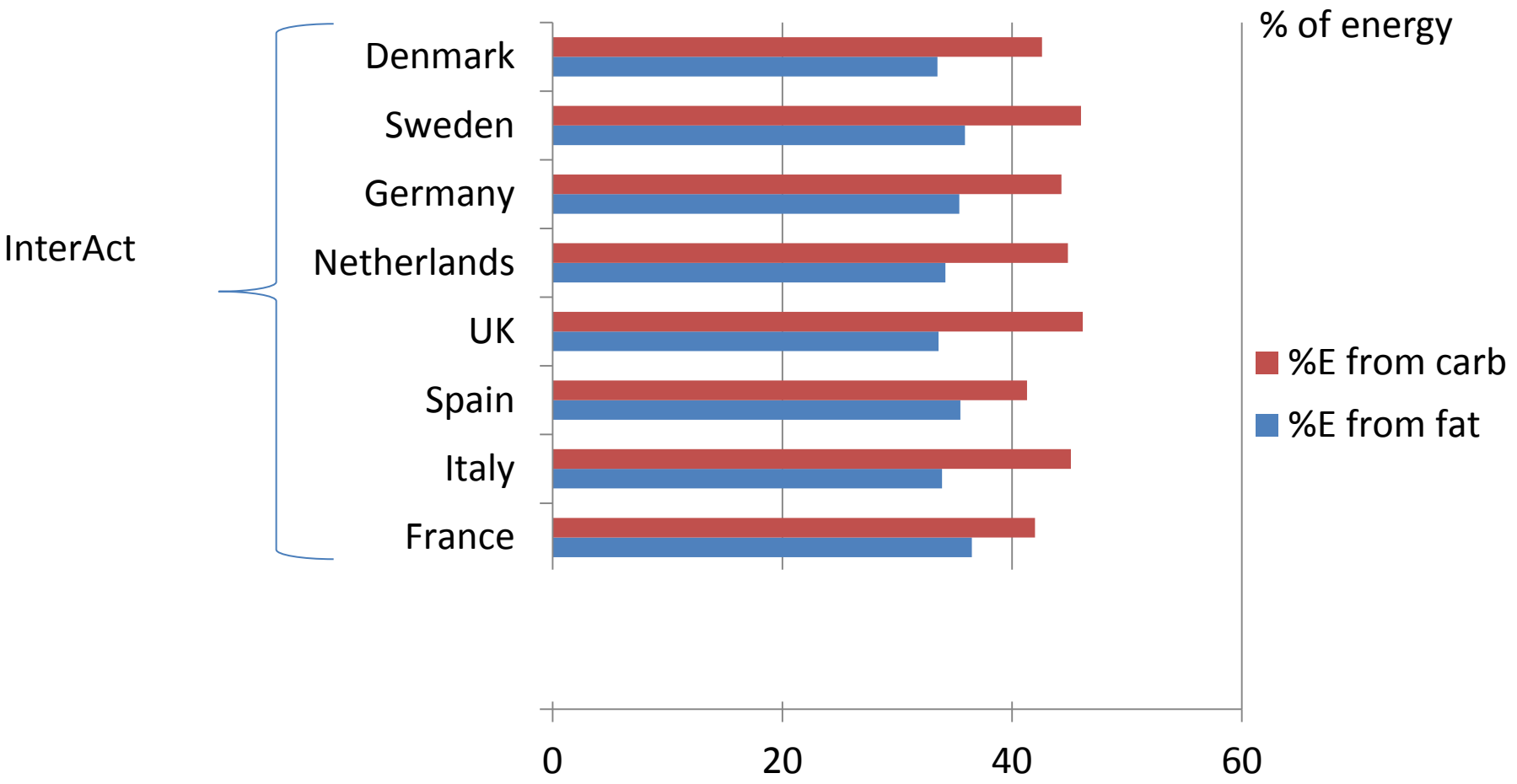
Global variation in carbohydrate intake



Between-population variance in lifestyle exceeds that within populations

Source: FAO Statistics Division 2010

Percentage energy (%E) from fat and carbohydrates



Source: Nanri et al, Am J Clin Nutr, 2011

How to realise the vision of bringing data together to allow the study of between-population differences in risk

- 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

InterConnect

Global data for diabetes and obesity research

Acknowledgement

- This project is funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 602068.

Connect with us

- InterConnect@mrc-epid.cam.ac.uk
- www.interconnect-diabetes.eu

Programme of the day

- Session 1 – Setting the scene
- **Session 2 – Challenges of current data sharing models**
- Session 3 – Vision of a changed paradigm
- Session 4 – Next steps – what can we do to move towards this changed paradigm

Session 1: Challenges of current data sharing models

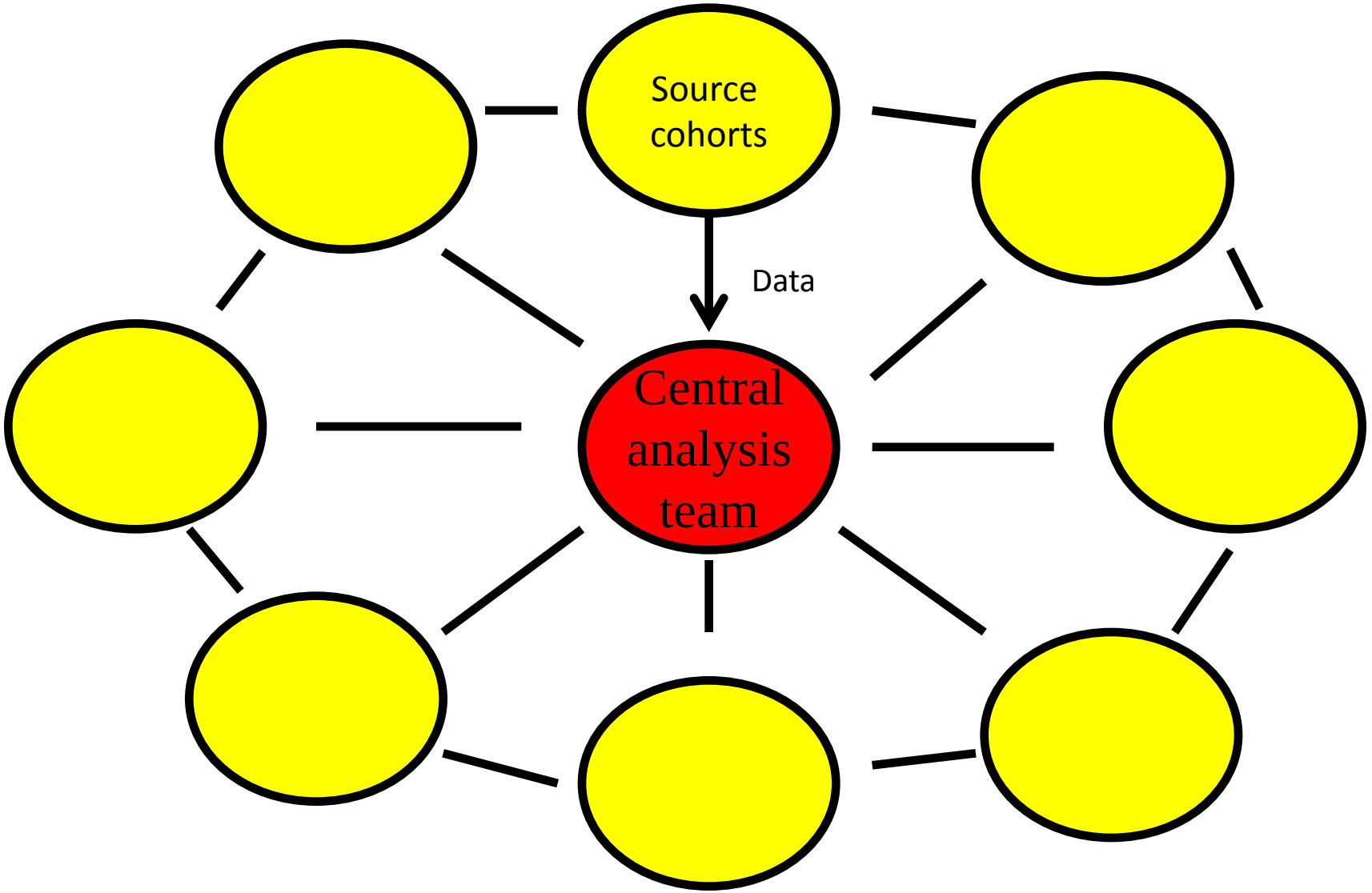
Present four alternative models about how data might be shared

Discuss models on your tables and work through the possible benefits and difficulties of each model

Think about issues from different perspectives – i.e that of a researcher, a funder etc

Try to think of a future world in which we are trying to collect multiple studies together across different countries

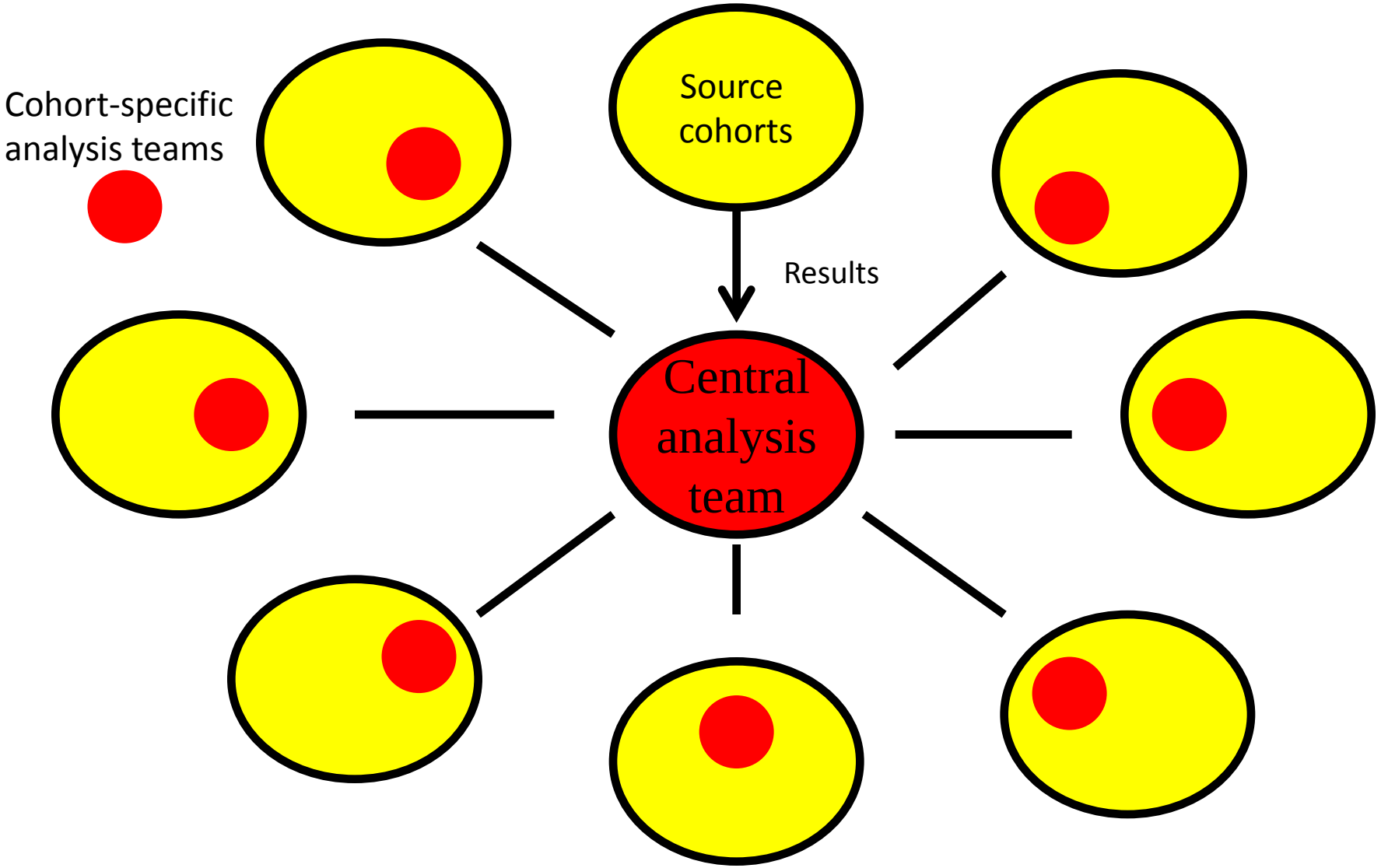
Sharing of data between cohorts using traditional collaboration/consortia agreements



Possible issues

- Considerable transactional burden
- Burden will increase exponentially as number of partners in consortia increases
- 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 as opposed to contribute

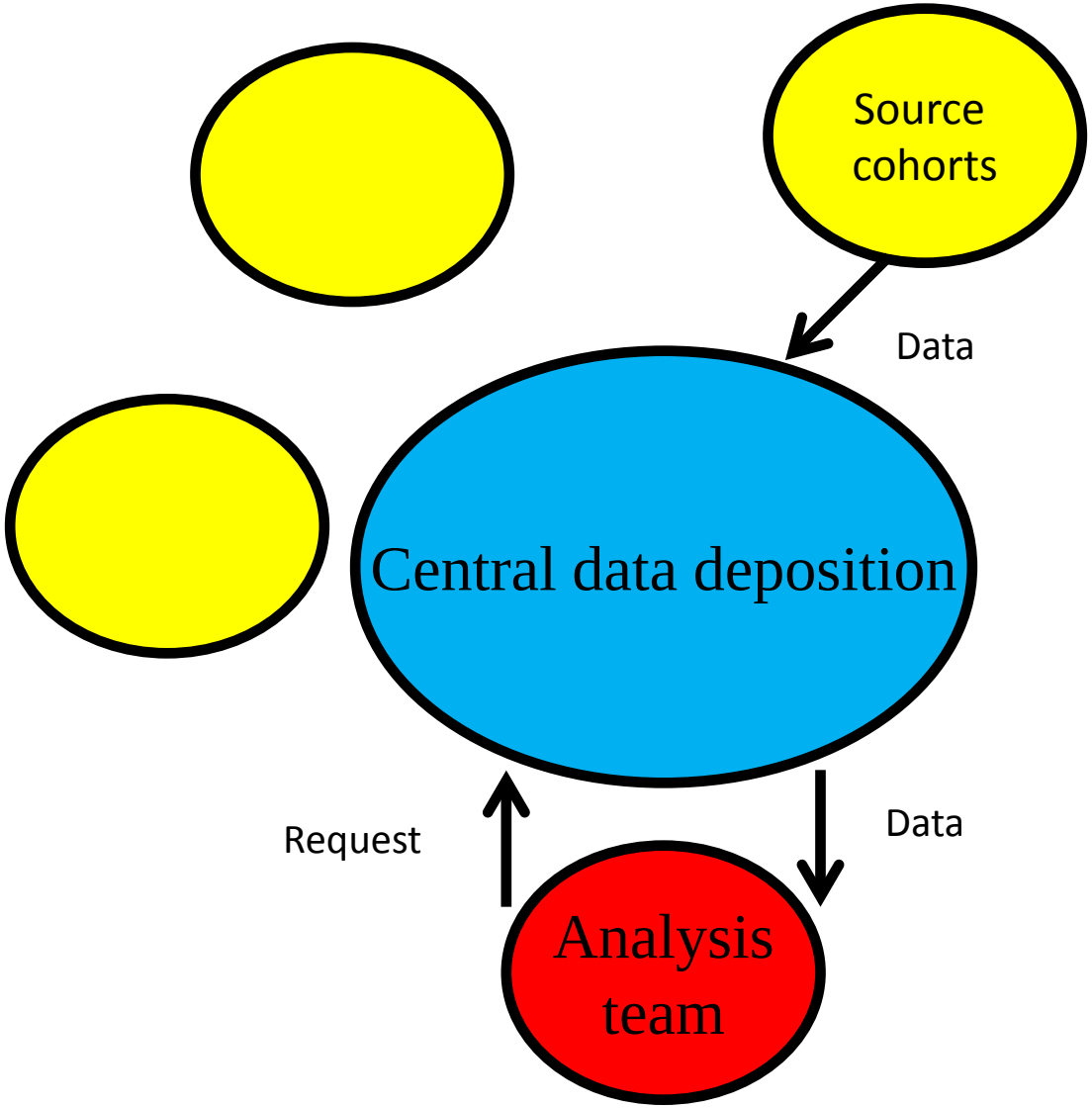
Ad hoc consortia - sharing of results



Possible issues

- Ad hoc consortia work well for genetic analyses, allowing sharing of RESULTS without administrative or organisational complexity
- Limits of meta-analysing interaction terms from individual studies
- Difficulties of data harmonisation given limited attention
- Analysis is potentially missing major between-cohort variation
- Analytical effort is decentralised to individual studies who spend a massive amount of time servicing the work of others

Central deposition of data



Possible issues

- Approach works within some countries for some forms of data
- Likelihood of success for between-country collaboration low
- Unlikely to work for more complex forms of data
- Major governance, ethical and legal challenges
- Difficult to mandate for historical data

Federated meta-analysis

- Data stays within governance structure of source cohort
- Cohorts focus efforts on preparation of data and IT infrastructure for sharing
- Analytical effort more focused on the scientific – led questions

Programme of the day

- Session 1 – Setting the scene
- Session 2 – Challenges of current data sharing models
- **Session 3 – Vision of a changed paradigm**
- Session 4 – Next steps – what can we do to move towards this changed paradigm

InterConnect

Global data for diabetes and obesity research

Acknowledgement

- This project is funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 602068.

Connect with us

- InterConnect@mrc-epid.cam.ac.uk
- www.interconnect-diabetes.eu

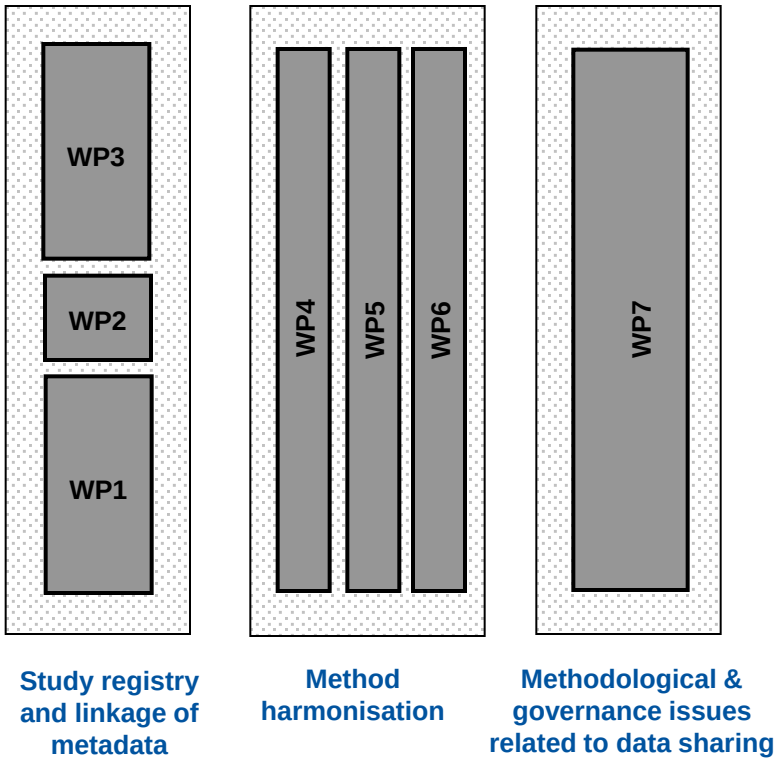
Programme of the day

- Session 1 – Setting the scene
- Session 2 – Challenges of current data sharing models
- **Session 3 – Vision of a changed paradigm**
- Session 4 – Next steps – what can we do to move towards this changed paradigm

The InterConnect Project

Application of existing tools

Core components of InterConnect



InterConnect

Global data for diabetes and obesity research

Study Registry

Matthias Schulze
German Institute of Human Nutrition Potsdam-Rehbrücke
Oct 10th 2014

This project is funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 602068.

Why a study registry is needed

- Researchers need to know what studies are being conducted
  - What resources are available globally
  - What study design was employed
  - What populations were recruited
  - Whether samples were stored
  - What data is available

Developing a study registry

- Tasks of the InterConnect project
  - Setup a database to include information about studies
  - Prepare a standardised web-based procedure for data input for project partners and external investigators
  - Prepare a registry website which hosts the visualization of the registry database

The InterConnect study registry

- a 2-phase registry
  - Phase 1: “broad and shallow”
    - Simple but useful information
    - Largely collected based on available/public information
  - Phase 2: in depth information
    - To be collected directly from studies

The InterConnect study registry

- Phase 1 information
  - General information (study name, contact persons, web link)
  - Study design
  - Ethnicity and race
  - Sampling frame
  - Recruitment target
  - Basic participant characteristics

The InterConnect study registry

- Phase 1 information
  - WPs will systematically review literature and extract study information
  - Cross-checked by study investigators

➢ Little burden for individual studies

➢ Large number of studies with basic information

How to create interest of studies to be included in registry ?

The InterConnect study registry

- Phase 2 information
  - To be collected from studies
  - Meta-data about available data
    - Data sources
    - Categories of available data (e.g. health, sociodemographic, lifestyle, physiological, biochemical, genotype information)

How to create commitment of studies to provide information?

The InterConnect study registry

- Long-term perspective
  - Keeping the registry up-to date
    - Inclusion of new studies
    - Update of data collection events, genotyping etc.
  - Sustainability of registry infrastructure
  - Promotion of its use by investigators

InterConnect

Global data for diabetes and obesity research

Data harmonisation

Nita Forouhi
MRC Epidemiology Unit, University of Cambridge
10th October 2014

This project is funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 602068.

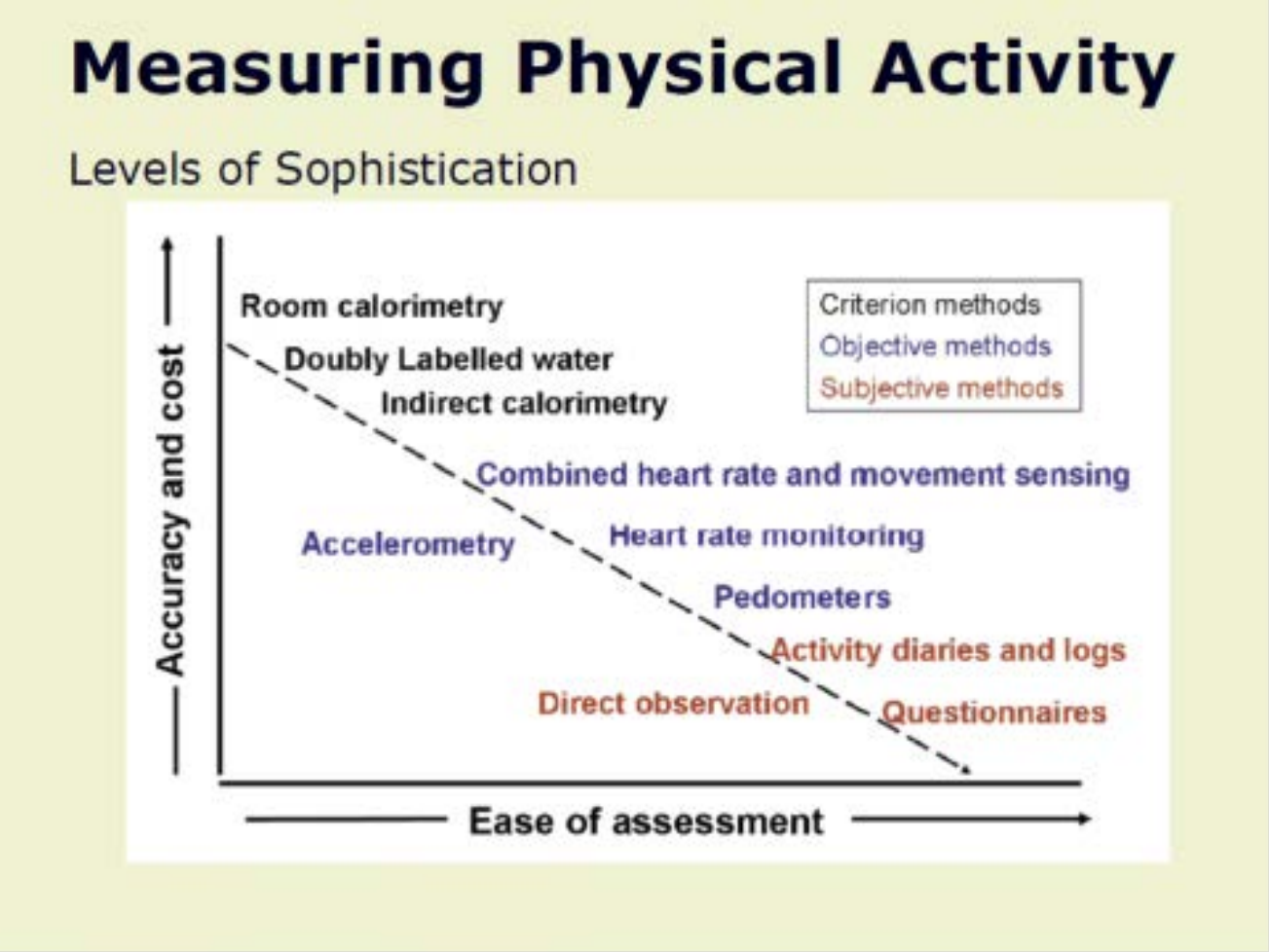
Data harmonisation: what is it?

- Data harmonisation
- 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 recoding or modifying variables so that they are comparable across research studies
- enables the synthesis of primary data from studies

Data harmonisation: context

- In the health context, in broad terms, data are collected for variables of “exposures” and “outcomes”
- Exposures and outcomes can be assessed using different measurement methods
- Different methods are used for a variety of reasons
  - What is known to the researcher
  - What is pragmatic, feasible, affordable
  - What tools are available for data collection
  - What tools are available for analysis

Example: choice of measurement



Variation in questionnaires: physical activity

Q're Name	Timeframe	Domains of activity
CARDIA physical activity history		
EPIC Physical Activity Questionnaire (**EPAQ, EPAQ2**)	Last 12 months	
Framingham Physical activity index		
Historical leisure activity questionnaire		
International Physical Activity Questionnaire (**IPAQ**)	Last 7 days	
Paffenbarger Physical Activity Questionnaire		
Recent Physical Activity Questionnaire (**RPAQ**)	Last month	
Stanford Usual Activity Questionnaire		
Tecumseh Occupational PAQ		
WHO Global Physical Activity Questionnaire (**GPAQ**)	Typical week	

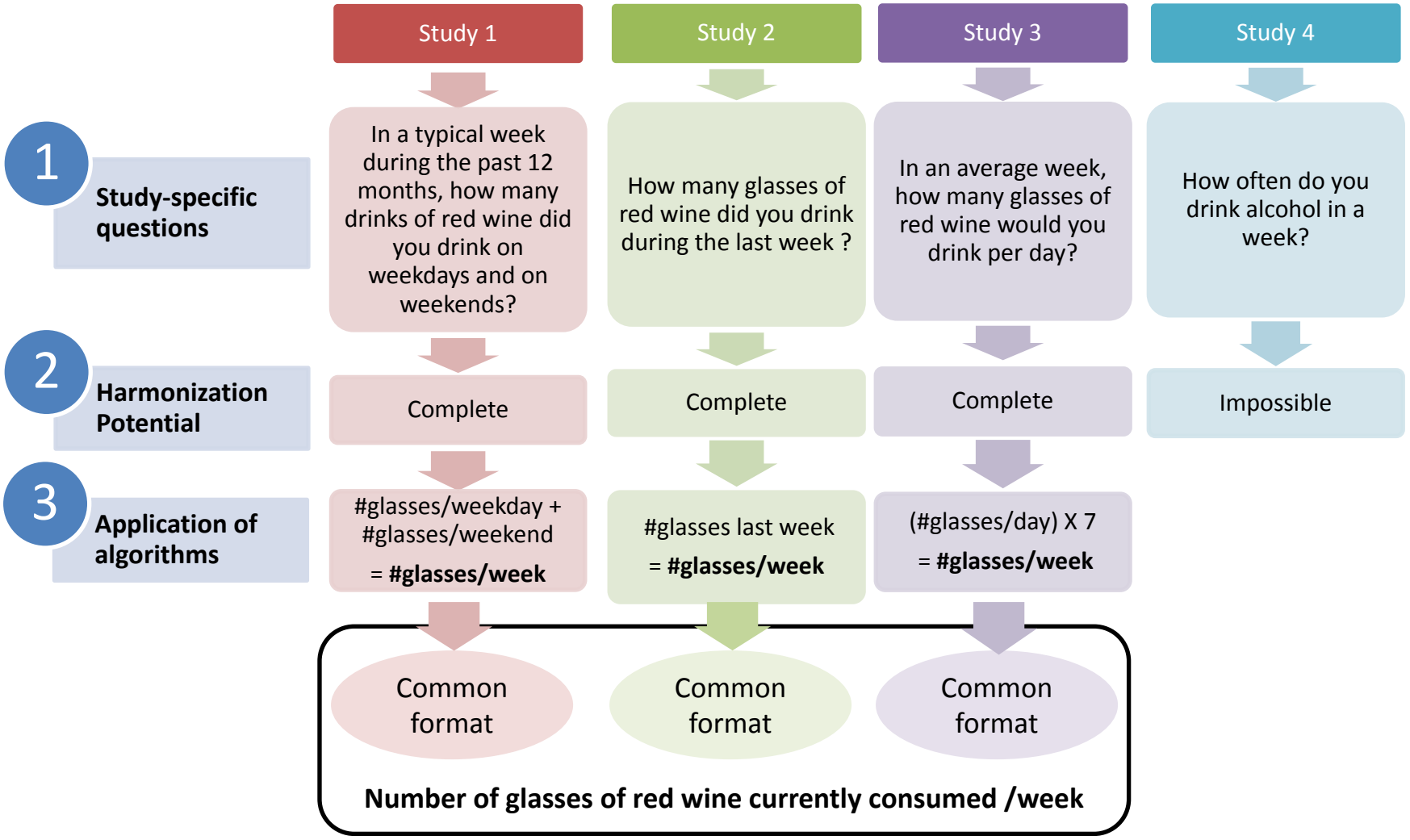
Review: Helmerhorst et al, IJBNPA 2012; 130 PA questionnaires included

Data harmonisation approaches

- For retrospectively collected data
  - Have to work with available variables
  - Work with study registry for list of available studies and related meta-data to assess harmonisation potential
  - Catalogue a listing of variables of interest
- For prospective data
  - Can define the optimum way to collect data across studies

Data harmonisation and processing: A lifestyle exposure

Number of glasses of red wine currently consumed/week



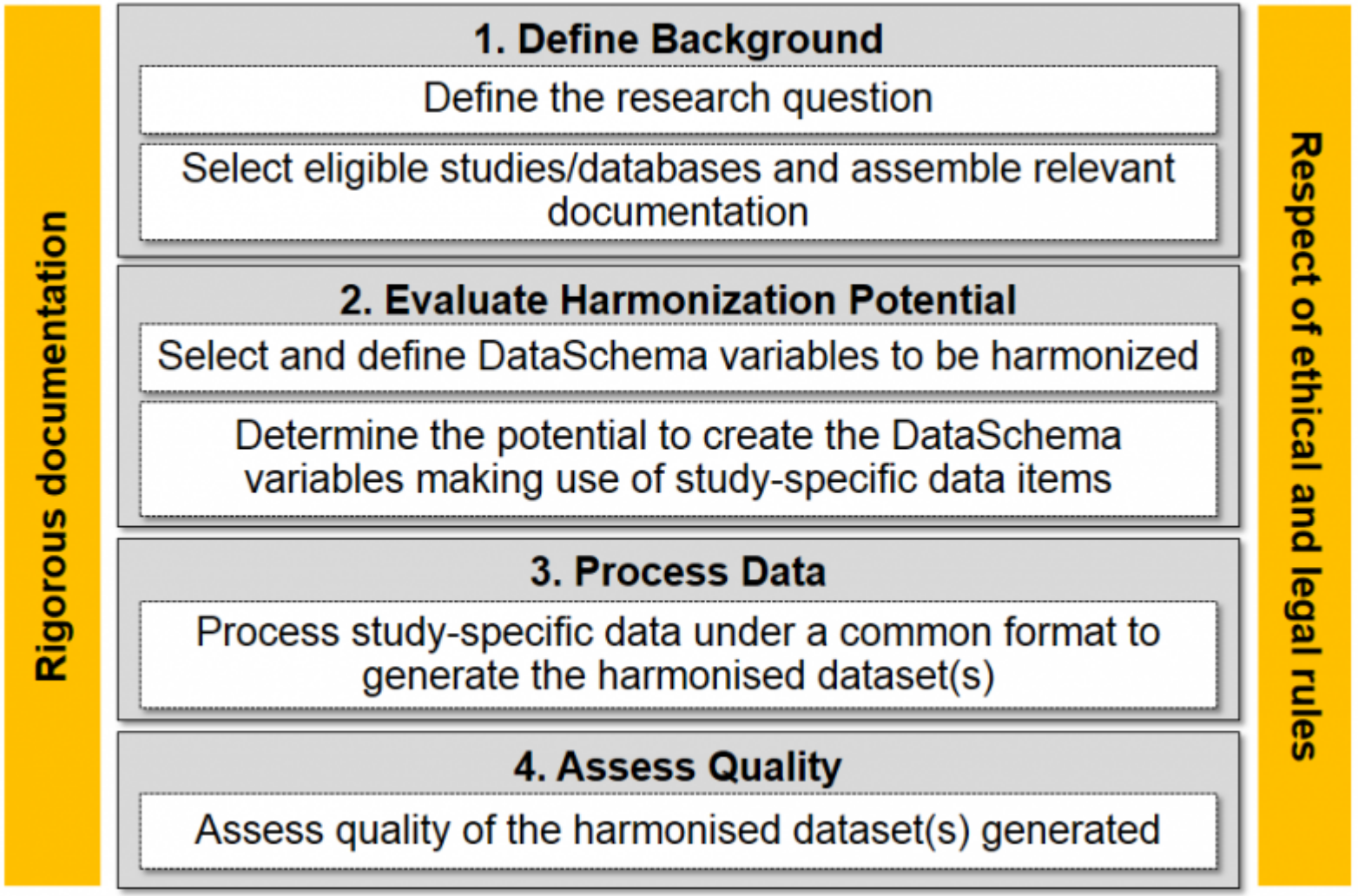
Slide from Isabel Fortier

Harmonisation: An outcome measure - fasting glucose concentration

	Study 1	Study 2	Study 3	Study 4
Glucose (mmol/l)	√	× Measured in mg/dl	× HbA1c	×
Glucose (mmol/l)	√	Apply conversion factor	Derive from HbA1c by applying conversion factor	×
Glucose (mmol/l)	√	√	√	×

Doiron D et al, 2013

Steps in data harmonisation

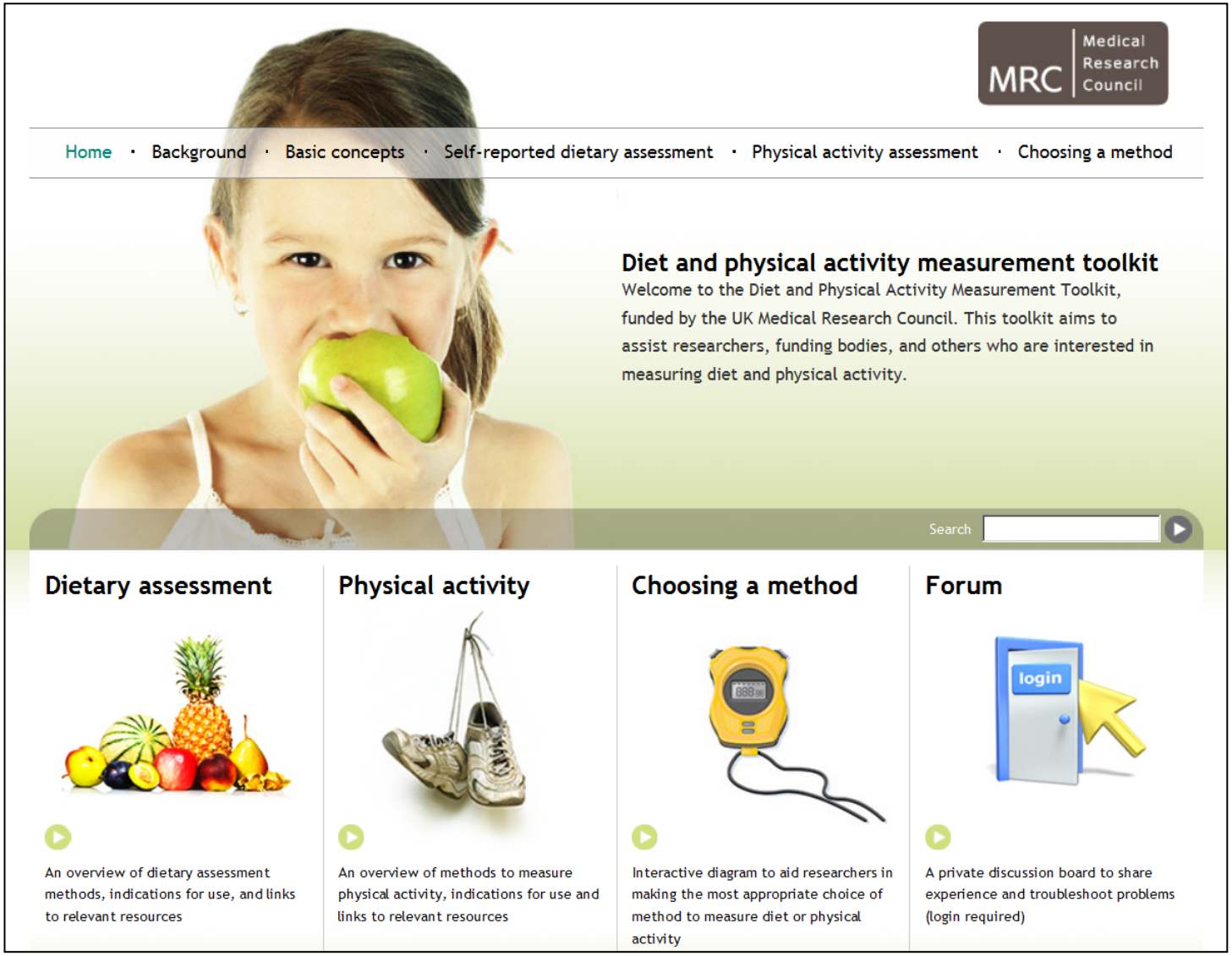


Prospective data harmonisation

- Greater comparability of future studies
- Can define and agree the optimum measures and procedures across studies
- Toolkit development to facilitate more unified approaches to data collection
  - Signpost researchers to methods that are fit for purpose

Diet and Physical Activity Assessment ("DAPA") Toolkit

www.dapa-toolkit.mrc.ac.uk



Population Health Sciences Measurement Toolkit (2014-15)

Population Health Sciences Measurement Toolkit

Home | Basic Concepts | Dietary Intake | Physical Activity | Anthropometry | Tobacco Use | Alcohol Consumption | Glossary Of Terms

MRC Population Health Sciences Research Network

Measure using methods fit-for-purpose.

The MRC Population Health Sciences Measurements Toolkit is a free, 'one-stop' resource to help you identify the most robust and up-to-date methods for the assessment of diet, physical activity, anthropometry, tobacco use and alcohol consumption.

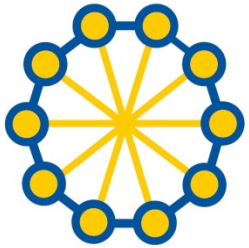
This website is designed for researchers planning to measure these lifestyle factors in groups or populations, providing them with targeted methodological guidance to choose a method that is most appropriate, valid, and cost-effective for the purposes of their study. Pick a fruit from the tree to explore your domain of interest....

New toolkit: the end product

- **General update/revision**
  - Incorporate recent technical developments
  - Include objective measurement, particularly for dietary intake
  - “Future-proofing” the Toolkit: re-design structure to make incremental updates easier to implement in future
- **Expansion of scope**
  - Include anthropometry, smoking and alcohol consumption, in addition to original diet and physical activity measures
- **Improved ‘decision matrix’ and enable access to methods**
  - dynamic process that responds iteratively to user study profile and objectives, producing a range of tailored suggestions
  - help to access these methods through signposting sources
- **More accessible web content**
  - greater use of multimedia learning resources and signposting links

Harmonisation- Summary

- Harmonisation is about optimising data for re-use in combined analyses across multiple studies
- We will harmonise methods for self-reported exposures across existing studies
- We will harmonise methods for objectively measured exposures across existing studies
- We will develop an online tool for signposting researchers to relevant methods for measurement of key exposures
- The same approaches apply to harmonisation of exposures and outcomes
- Harmonisation also applies to analytical approaches



InterConnect

Global data for diabetes and obesity research

Federated Analysis

Isabel Fortier
McGill University (Canada)
October 10th 2014

This project is funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 602068.

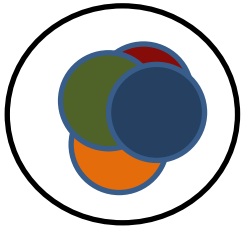
Analysing harmonized data

Summary data meta-analysis

- Study-specific data analysis (independent analyses followed by a meta-analysis combining the study-level estimates)

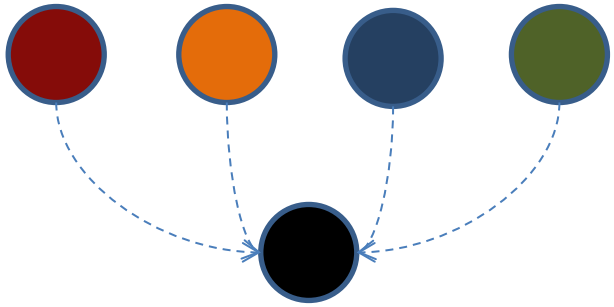
Pooled analysis

- Pooled data analysis (data transferred to a central server and pooled to be analyzed)



Federated analysis

- Federated data analysis (centralized analysis, but the individual-level participant data remain on local servers)



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.

Biobank Standardization and Harmonization for Research Excellence in the European Union

- European Union FP7-funded project (2012-2016)
- Lead-PI Ronald Stolk (The Netherlands)
- Mission: Ensure the development of harmonized measures and computing infrastructures across biobanks in Europe

Multiple scientific questions; long term harmonization agenda

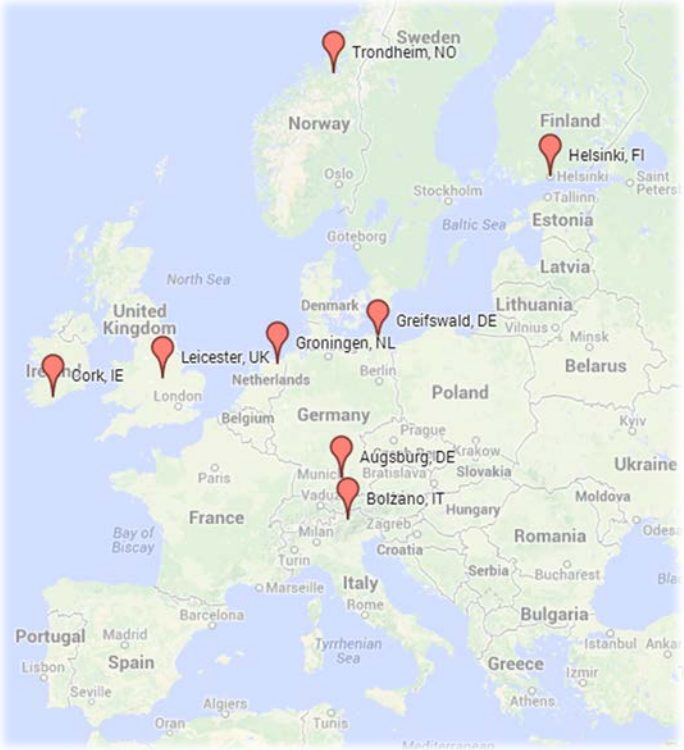
Federated data infrastructure

BioSHaRE's Healthy Obese Project (B Wolffenbuttel)

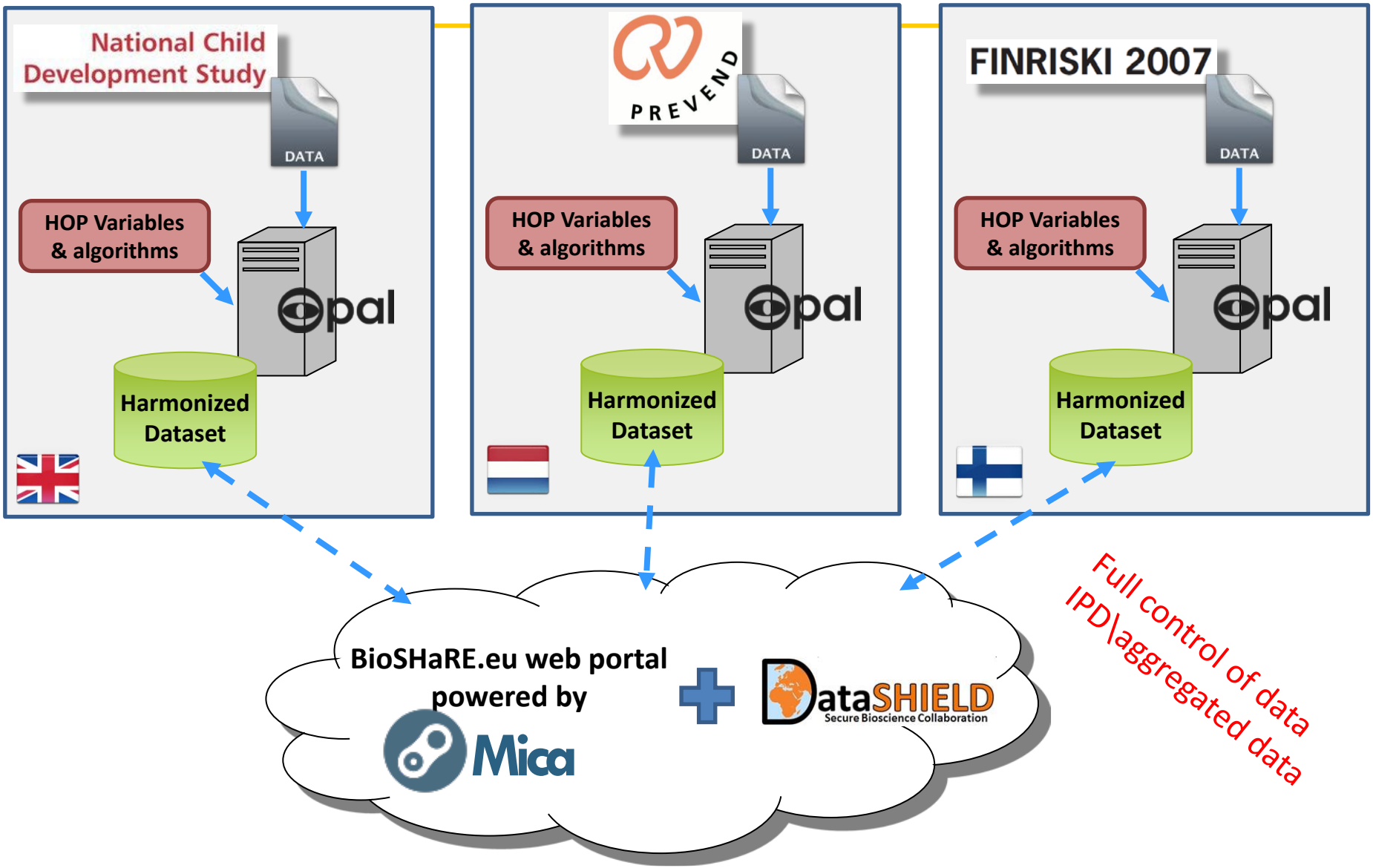
Aims:

- Evaluate the prevalence of the metabolically healthy obese
- Assess lifestyle determinants of healthy obesity
- Explore genetic determinants and metabolic profiling related to healthy obesity

- 10 studies, 7 countries
- ~ 200,000 participants

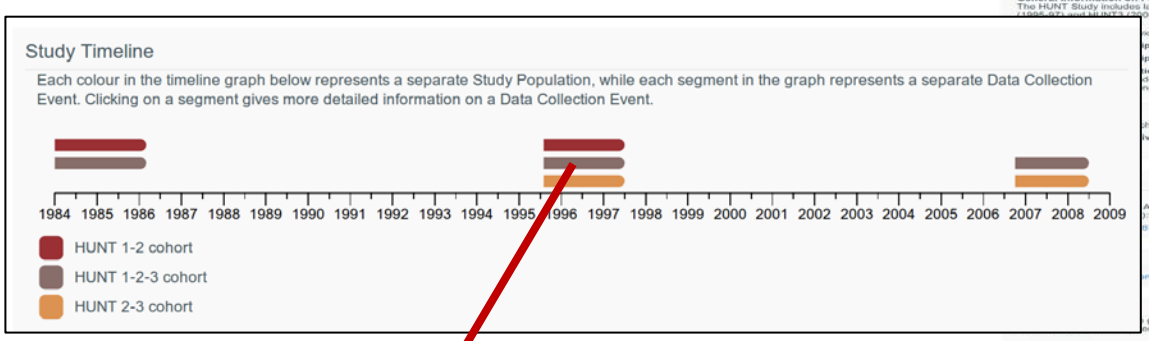


10 Harmonized Datasets Using Opal



Gather knowledge

Document study design, methods and contents



HUNT2

Questionnaires (eight) with self-reported health, quality of life, illness, diseases (several), major behavioural risk factors and socio-economic position. Disease-specific questionnaires for hypertension, diabetes and lung diseases. Anthropometric measures (weight, height, waist and hip circumference), blood pressure and heart rate, spirometry, forearm bone mineral density (BMD), and vision. Venous blood samples, analysed cholesterol (total and HDL), triglycerides, glucose, Se-Fe, transferrin and creatinine, thyroid-stimulationg hormone (TSH), calcium, parathyroid hormone (PTH) and stored (serum aliquots stored at –80°C), DNA extracted. Urine: microalbumine and creatinine (sub-groups).

The number of participants with biological samples was 65 195.

Start Year: 1995 (August)

End Year: 1997 (June)

Data sources: Questionnaires
Physical measures
Biological samples

Biological samples: Blood
Urine

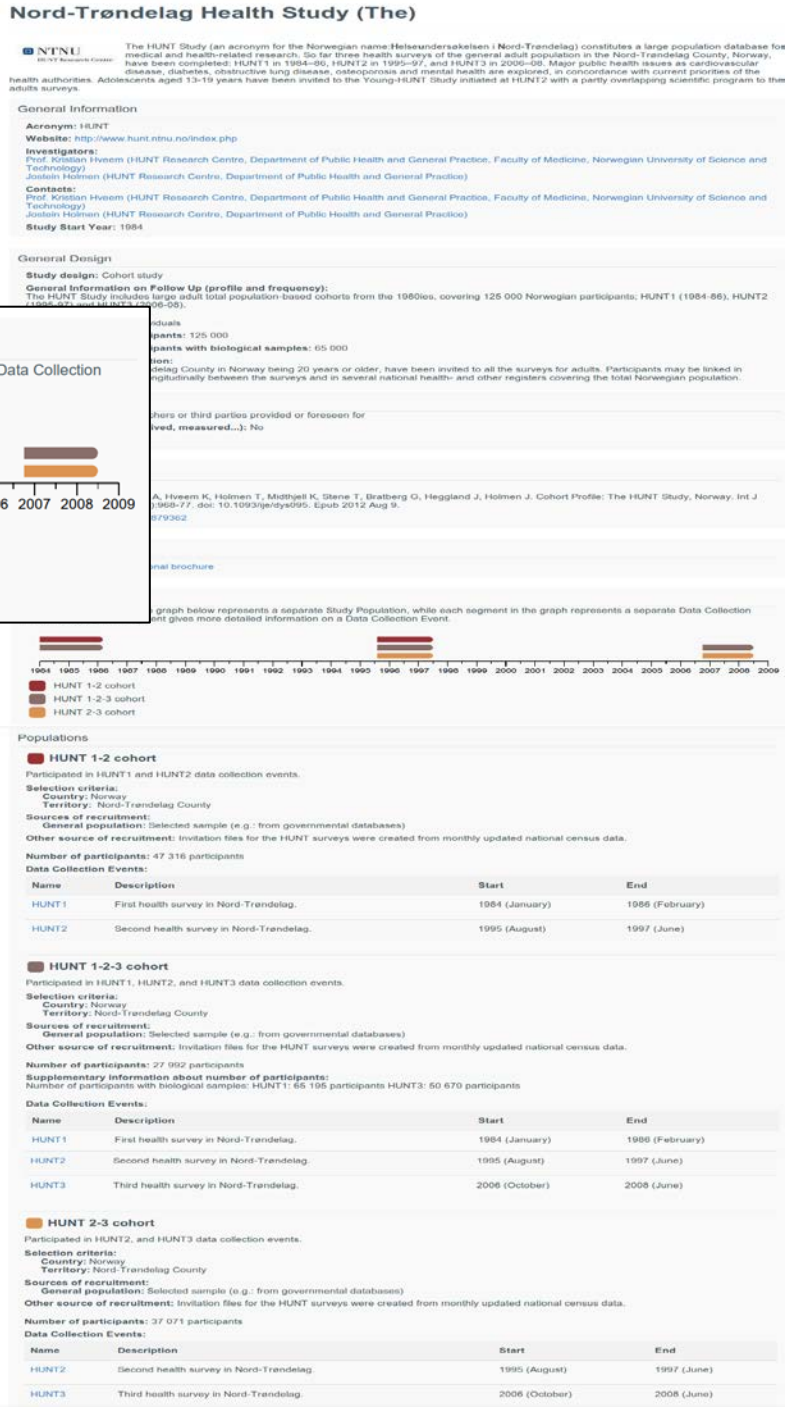
Population: HUNT 1-2 cohort
HUNT 1-2-3 cohort
HUNT 2-3 cohort

Datasets: HUNT2

Documents

Questionnaires: HUNT Questionnaire 1
HUNT Questionnaire 2 for men
HUNT Questionnaire 2 for women
HUNT2 Diabetes Questionnaire

Searchable dataset



Nord-Trøndelag Health Study (The)

NTNU

The HUNT Study (an acronym for the Norwegian name Helseundersøkelsen i Nord-Trøndelag) constitutes a large population database for medical and health-related research. So far three health surveys of the general adult population in the Nord-Trøndelag County, Norway, have been completed: HUNT1 in 1984–86, HUNT2 in 1995–97, and HUNT3 in 2006–08. Major public health issues as cardiovascular disease, diabetes, obstructive lung disease, osteoporosis and mental health are explored, in concordance with current priorities of the health authorities. Adolescents aged 13-19 years have been invited to the Young-HUNT Study initiated at HUNT2 with a partly overlapping scientific program to the adults surveys.

General Information

Acronym: HUNT
Website: http://www.hunt.ntnu.no/index.php
Investigators:
Prof. Kristian Hveem (HUNT Research Centre, Department of Public Health and General Practice, Faculty of Medicine, Norwegian University of Science and Technology)
Jostein Holmen (HUNT Research Centre, Department of Public Health and General Practice)
Contacts:
Prof. Kristian Hveem (HUNT Research Centre, Department of Public Health and General Practice, Faculty of Medicine, Norwegian University of Science and Technology)
Jostein Holmen (HUNT Research Centre, Department of Public Health and General Practice)
Study Start Year: 1984

General Design

Study design: Cohort study
General Information on Follow Up (profile and frequency):
The HUNT Study includes large adult total population-based cohorts from the 1980ies, covering 125 000 Norwegian participants; HUNT1 (1984-86), HUNT2 (1995-97) and HUNT3 (2006-08).

...iduals
...pants: 125 000
...pants with biological samples: 65 000
...ion:
...delag County in Norway being 20 years or older, have been invited to all the surveys for adults. Participants may be linked in ...ngitudinally between the surveys and in several national health- and other registers covering the total Norwegian population.

...hers or third parties provided or foreseen for
...ived, measured...): No

...A, Hveem K, Holmen T, Midthjell K, Stene T, Bratberg G, Heggland J, Holmen J. Cohort Profile: The HUNT Study, Norway. Int J ...968-77. doi: 10.1093/ije/dyr095. Epub 2012 Aug 9.
...879362

...nal brochure

...graph below represents a separate Study Population, while each segment in the graph represents a separate Data Collection ...ent gives more detailed information on a Data Collection Event.

1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
HUNT 1-2 cohort
HUNT 1-2-3 cohort
HUNT 2-3 cohort

Populations

HUNT 1-2 cohort

Participated in HUNT1 and HUNT2 data collection events.

Selection criteria:
Country: Norway
Territory: Nord-Trøndelag County
Sources of recruitment:
General population: Selected sample (e.g.: from governmental databases)
Other source of recruitment: Invitation files for the HUNT surveys were created from monthly updated national census data.
Number of participants: 47 316 participants
Data Collection Events:

Name	Description	Start	End
HUNT1	First health survey in Nord-Trøndelag.	1984 (January)	1986 (February)
HUNT2	Second health survey in Nord-Trøndelag.	1995 (August)	1997 (June)

HUNT 1-2-3 cohort

Participated in HUNT1, HUNT2, and HUNT3 data collection events.

Selection criteria:
Country: Norway
Territory: Nord-Trøndelag County
Sources of recruitment:
General population: Selected sample (e.g.: from governmental databases)
Other source of recruitment: Invitation files for the HUNT surveys were created from monthly updated national census data.
Number of participants: 27 992 participants
Supplementary information about number of participants:
Number of participants with biological samples: HUNT1: 65 195 participants HUNT3: 50 670 participants
Data Collection Events:

Name	Description	Start	End
HUNT1	First health survey in Nord-Trøndelag.	1984 (January)	1986 (February)
HUNT2	Second health survey in Nord-Trøndelag.	1995 (August)	1997 (June)
HUNT3	Third health survey in Nord-Trøndelag.	2006 (October)	2008 (June)

HUNT 2-3 cohort

Participated in HUNT2, and HUNT3 data collection events.

Selection criteria:
Country: Norway
Territory: Nord-Trøndelag County
Sources of recruitment:
General population: Selected sample (e.g.: from governmental databases)
Other source of recruitment: Invitation files for the HUNT surveys were created from monthly updated national census data.
Number of participants: 37 071 participants
Data Collection Events:

Name	Description	Start	End
HUNT2	Second health survey in Nord-Trøndelag.	1995 (August)	1997 (June)
HUNT3	Third health survey in Nord-Trøndelag.	2006 (October)	2008 (June)

Explore information available across studies/data collection events

Search by domains - terms :

New version available in Novembre

	Quebec Longitudinal St...				FRelE: Fragilité, une ...		
	T1	T2	T4	T3	T0	T1	T2
Life habits/Behaviours							
Tobacco use	5	2	2	2	6	6	6
Alcohol use	20	12	13	12	9	9	9
Illicit drug use	-	-	-	-	-	-	-
Nutrition	560	86	151	86	8	8	8
Food intake and frequency	449	21	89	24	-	-	-
Milk products	40	1	8	1			
Meat, eggs, fish and alternatives	30	5	10				
Vegetables and fruits	130	2	27				
Cereals, bread and starches	60	1	12				
Sweet and baked goods	10	1	3				

- MEDICAL HEALTH INTERVENTIONS/HEALTH SERVICES UTILIZATION
- MEDICATION
- REPRODUCTIVE HEALTH AND HISTORY
- PARTICIPANT'S EARLY LIFE/CHILDHOOD
- LIFE HABITS/BEHAVIOURS
- SOCIODEMOGRAPHIC/SOCIOECONOMIC CHARACTERISTICS
- PHYSICAL ENVIRONMENT
- SOCIAL ENVIRONMENT
- PERCEPTION OF HEALTH/QUALITY OF LIFE
- ANTHROPOMETRIC STRUCTURES
- BODY STRUCTURES
- BODY FUNCTIONS
- LABORATORY MEASURES

Study-specific variable

Q128A

Description

Label:
Current use of alcohol

Dataset:
Finrisk 2007 Study Dataset

Value Type:
Text

Unit:

Repeatable:
No

Categories

Name	Label	Missing
1	Yes, at least once a month	
2	Yes, more seldom than once a month	
3	No, because I quit using alcohol...years ago	
4	I have never used alcohol (proceed to question 120)	

Dimensions

Data Source:
Questionnaire

Individual Participant Measures:
Health Outcomes and Risk Factors Measures

Life Habits/Behaviours:
Alcohol Use

Essence:
Occurrence/Status/Named category

Target:
Participant

Period:
Currently

Measure:
Categorical ordered

Statistics

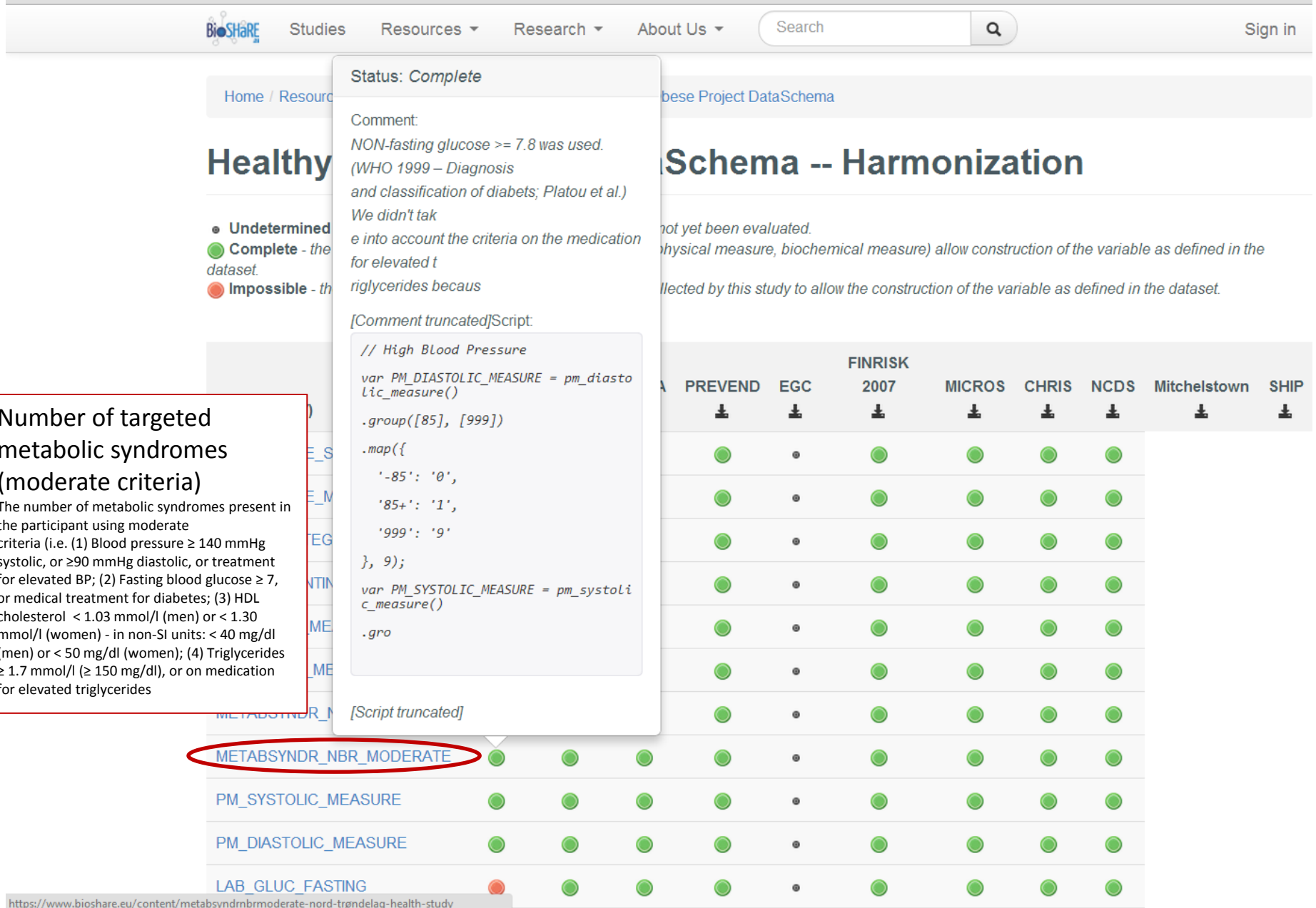
Value	National FINRISK Study 2007 (The)
1	771
2	733
3	739
4	757
All	3000


Create harmonized variables

DataSchema (core variables) = 97

Domain of information (# of variables)	e.g.
Diseases of the circulatory system (3)	History of Stroke, History of Hypertension
Endocrine, nutritional and metabolic diseases (2)	History of Diabetes, Type of Diabetes
Medication Intake (4)	Current Use of Antihypertensive Medication, Lipid Lowering Medication, Glucose Lowering Medication
Alcohol Use (6)	Current Use of Alcohol, Current Quantity of Beer Consumed
Tobacco Use (7)	Current Tobacco Smoker, Current Quantity of Cigarettes Smoked
Food Intake and Frequency (23)	Current Consumption Frequency of Fruits, Current Consumption Frequency of Soft Drinks
Nutritional behaviours (7)	Currently Follows a Vegetarian Diet, Currently Follows a Diabetic Diet
Working Status (5)	Employment Status, Current Job Title (ISCO 88)
Education Level (6)	Number of Years of Education, Highest Level of Education
Household Status (4)	Net Household Income, Marital Status
Parity/Gravidity (1)	Number of Live Births Mothered
Gender / Age (4)	Gender, Age (continuous + categorical)
Residence / Birth Location (3)	Current Country of Residence, Country of Birth
Anthropometric structures / Body Function (6)	Height, Weight, Measured Systolic Blood Pressure,
Biochemical measures (9)	HDL Cholesterol, Glucose, Triglycerides, Inflammation Marker (hsCRP)
Constructed variables and others (7)	BMI, Healthy Obese, Number of Metabolic Syndromes, Year of Interview

Document centrally: Transparency, reproducibility, re-usage…

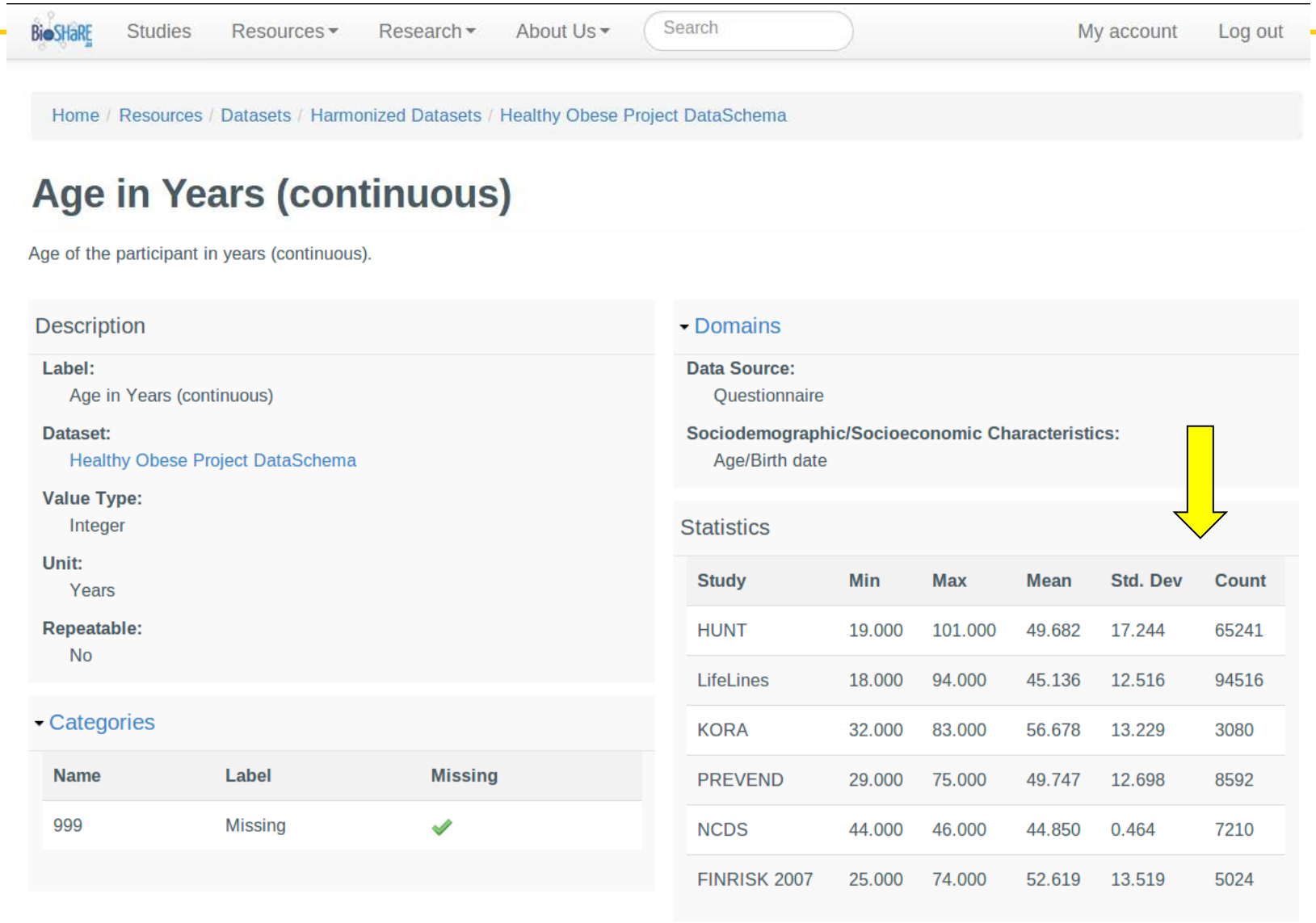


Number of targeted metabolic syndromes (moderate criteria)

The number of metabolic syndromes present in the participant using moderate criteria (i.e. (1) Blood pressure ≥ 140 mmHg systolic, or ≥90 mmHg diastolic, or treatment for elevated BP; (2) Fasting blood glucose ≥ 7, or medical treatment for diabetes; (3) HDL cholesterol < 1.03 mmol/l (men) or < 1.30 mmol/l (women) - in non-SI units: < 40 mg/dl (men) or < 50 mg/dl (women); (4) Triglycerides ≥ 1.7 mmol/l (≥ 150 mg/dl), or on medication for elevated triglycerides


Co-analyse harmonized data

Real time summary statistics on harmonized data



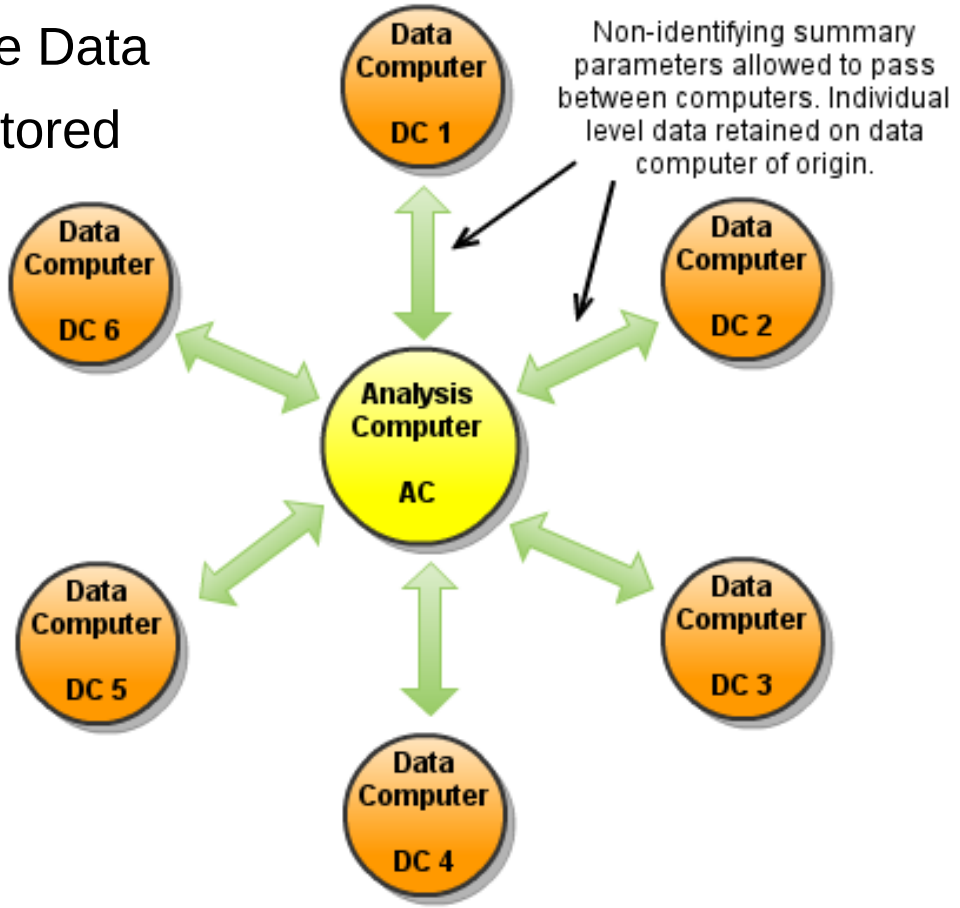
DataSHIELD
Secure Bioscience Collaboration

Federated analysis using DataSHIELD…

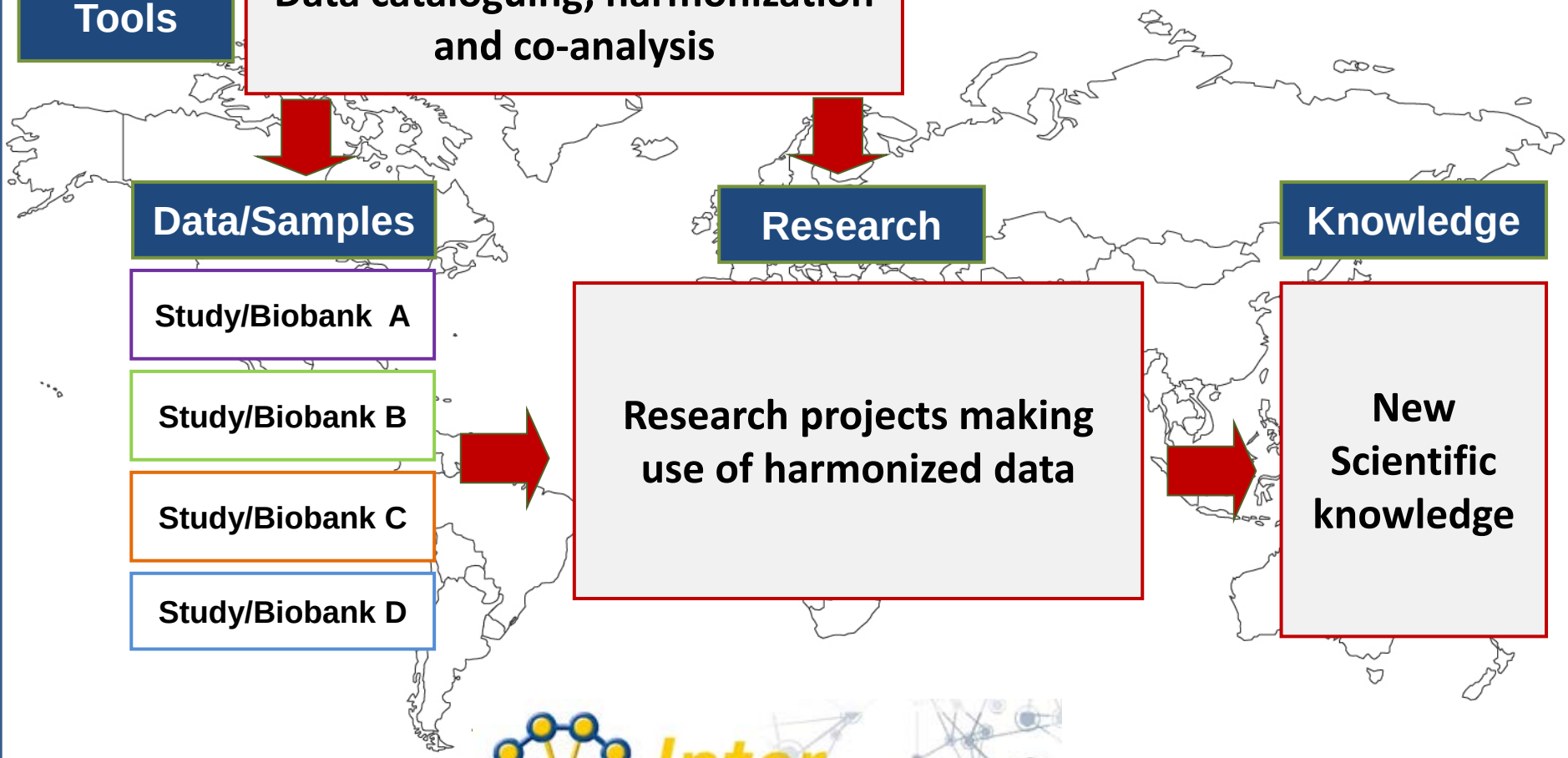
The Analysis Computer (AC) send iteratively requests for fitting a given GLM to the Data Computers (DC) on which data are stored

Only summary statistics are sent back to the AC after each iteration

Eventually, iterations converge to the **same** result as if the model was fitted directly to the pooled data.



Data harmonization/federation landscape



InterConnect

Maelstrom Research

- International research program (www.maelstrom-research.org)
- Created in 2012, but based on partnerships established since 2004 and leverage by the BioSHaRE.EU project
- Co-Investigators from Canada, The Netherlands and United Kingdom
- Collaboration with over 15 international networks and research partners
- Objectives: (1) Achieve methodological research; (2) generate software to support data cataloguing, harmonization, processing and integration; (3) create web-based catalogues and harmonization platform

InterConnect

Global data for diabetes and obesity research

Ethical, Legal, Social Issues

Ronald Stolk

Professor of Clinical Epidemiology

Coordinator BioSHaRE, co-PI Maelstrom Research Program

This project is funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 602068.

Ethical

- Federated data analysis deals with privacy issue at network level
  - Cohort PIs retain possession, control, and responsibility
  - Individual data stays within original cohorts, never shared
  - Analyses can be seen as equivalent to using anonymous data

Ethical

- Security of personal data systematically safeguarded
  - Impossible to externally access personal data—it never moves beyond cohort firewalls
  - Other safeguards: restricted coding, etc
- Proven methods, tools, systems
  - BioSHaRE, Maelstrom experience

Ethical

- No difference with traditional collaborative projects
- No additional institutional-level ELSI responsibility beyond original cohort
  - Data have been collected already
- Access regulations for each participating study
  - Compliance with participant consent, IRB review, data sharing committees, etc

Legal

- International Code of Conduct for Genomic and Health-Related Data Sharing
  - BioSHaRE in collaborating with P3G, the Global Alliance for Genomics and Health, IRDiRC (International Rare Diseases Research Consortium), H3Africa, and other organizations
- The Code promotes access to shared data, knowledge, and resources
- Ultimately, the Code will hopefully serve to promote data sharing and to sanction data misuse

Social

- Open access step further

- Covers researcher fear factor?
- Covers participant fear factor?

Addressing ELSI issues

- Changing ethical and regulatory perceptions
  - Data sharing ≠ security risk
  - Acceptance of a “new norm”
- Legal landscape shaped by "conventional" data sharing
  - Laws surrounding data (re-)usage and security vary
- Changing social and scientific views data sharing, collaboration
  - Encouraging and widening participation
  - Access/harmonization requires careful attention to metadata (data dictionary), English translation
  - Encourage collaboration, maximize local/small data resources
- Communication is key
  - Inform cohorts and IRBs: promote understanding, acceptance, endorsement
  - General population: promote trust in data sharing

ELSI support

- BBMRI-ERIC common service ELSI
- BioSHaRE newsletter July 2014
  - www.bioshare.eu -> about us
- BioSHaRE Tools roll-out meeting July 28, Milan, Italy
  - HandsOn Biobanks Conference 2015

InterConnect

Global data for diabetes and obesity research

Acknowledgement

- This project is funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 602068.

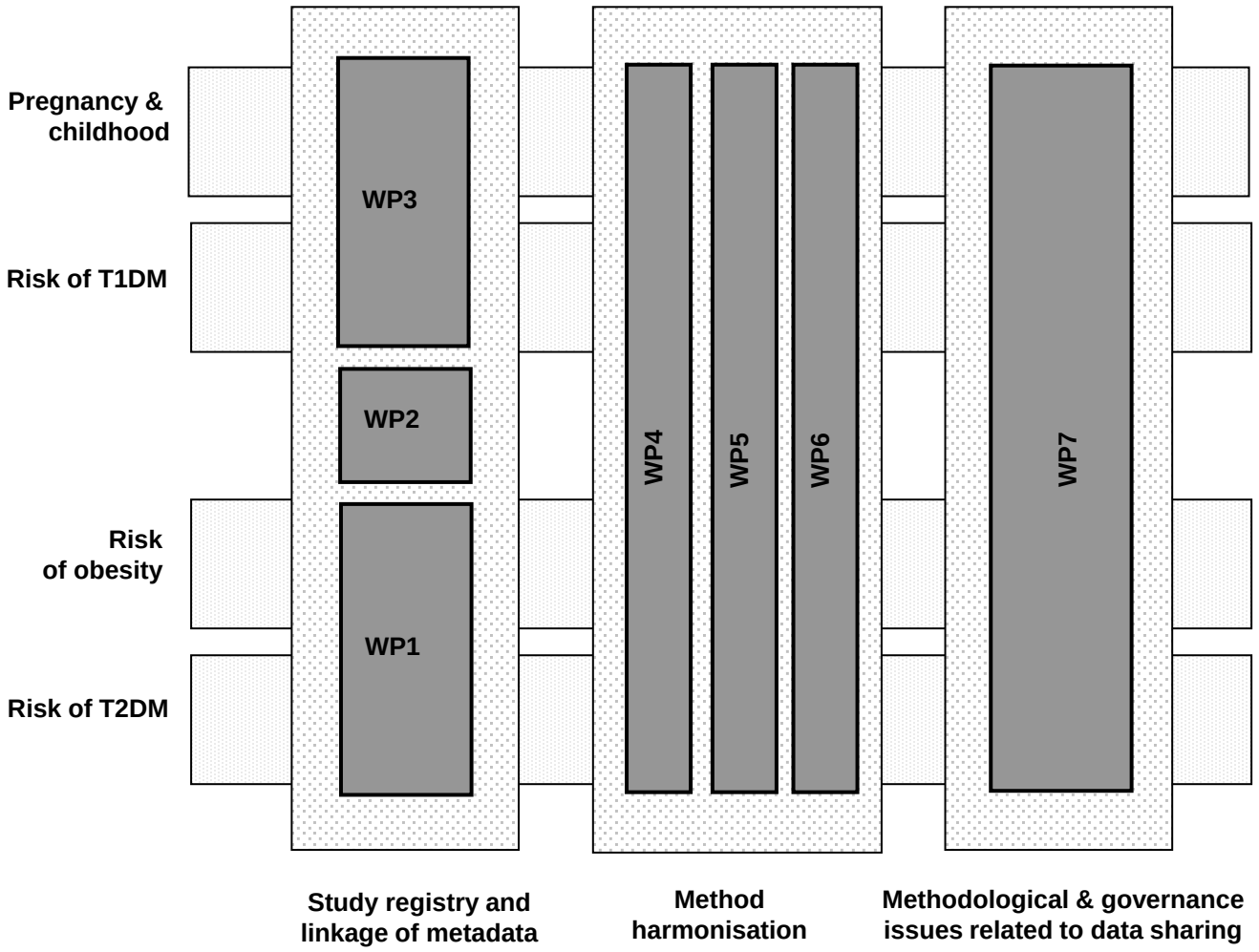
Connect with us

- InterConnect@mrc-epid.cam.ac.uk
- www.interconnect-diabetes.eu

Programme of the day

- Session 1 – Setting the scene
- Session 2 – Challenges of current data sharing models
- Session 3 – Vision of a changed paradigm
- **Session 4 – Next steps – what can we do to move towards this changed paradigm**

Engaging with researchers



- **Changing the paradigm →**

InterConnect

Global data for diabetes and obesity research

Session 4: Next Steps

Researcher engagement

Ruth Loos
Icahn School of Medicine at Mount Sinai, New York, NY
October 10, 2014

This project is funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 602068.

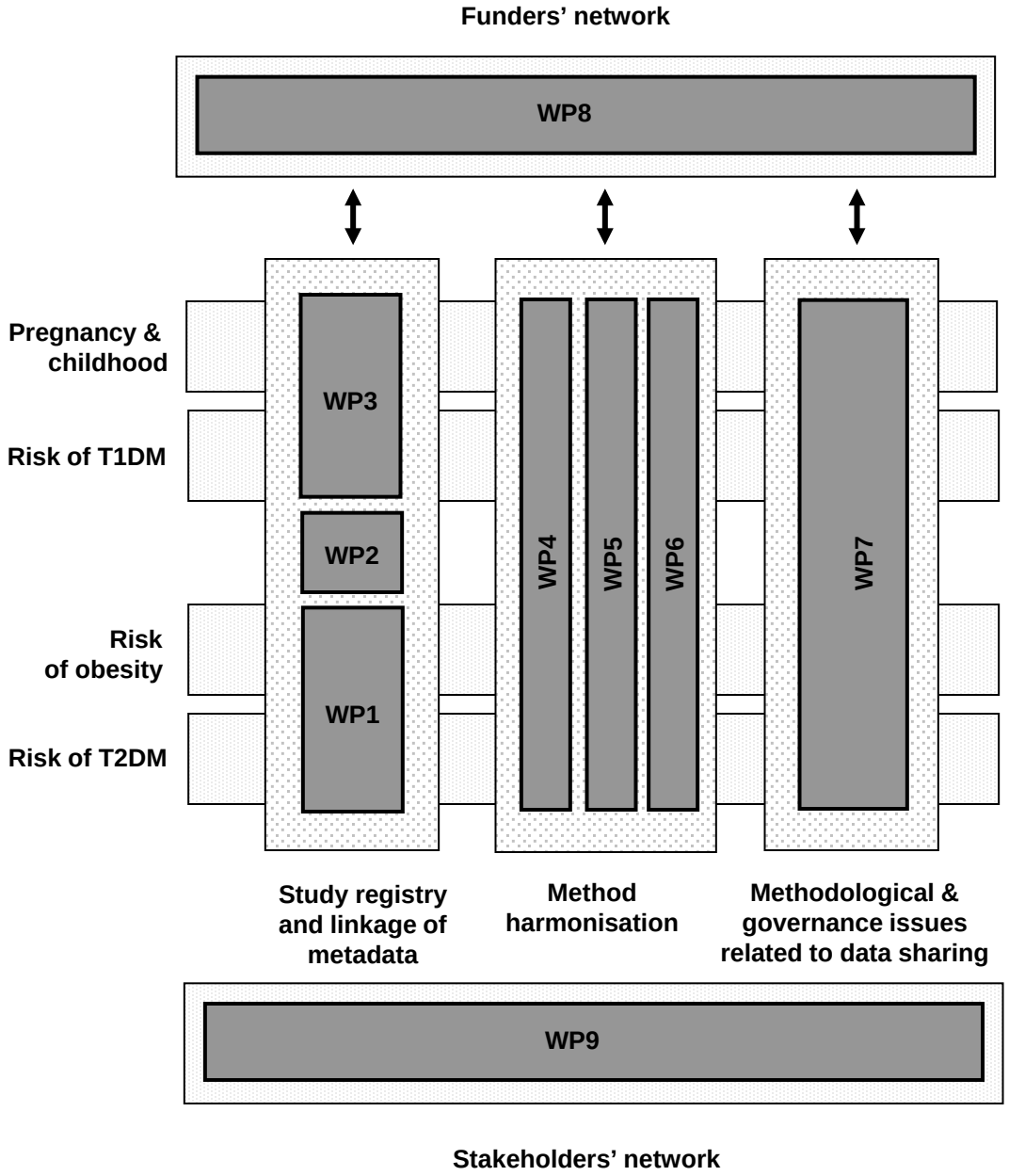
Sharing of data to date

1. Public sharing (e.g dbGap, EGA, …)
2. Sharing individual level genotype and phenotype data with collaborators (e.g Psychiatric Genomics Consortium)
3. Sharing summary statistics with collaborators (DIAGRAM+, MAGIC, AAGILE, MEDIA, GIANT, AAAGC, …)

Federated analyses

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

- **Changing the paradigm**
- **Engaging with researchers**
- **Engaging with funders and stakeholders**



An analogy for a paradigm shift in science

- The move towards open access publishing
- Requires an initial political will and an acceptance of a direction of travel
- Identification of barriers/obstacles to implementation
- Funder/researcher behaviour altered by some infrastructural changes and alterations in incentives

InterConnect

Global data for diabetes and obesity research

Acknowledgement

- This project is funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 602068.

Connect with us

- InterConnect@mrc-epid.cam.ac.uk
- www.interconnect-diabetes.eu