Global data for diabetes and obesity research

Global data for diabetes and obesity research

Host: Nick Wareham, InterConnect Co-ordinator & Director, MRC Epidemiology Unit, University of Cambridge, UK

Venue: pre-EASD Munich, 12 September 2016

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

Programme

14:30	**The InterConnect Project**
14:45	**Applying the InterConnect approach for federated meta-analysis** 1. Physical activity in pregnancy and neonatal anthropometric outcomes 2. Fish intake and risk of type 2 diabetes
16.00	**Future perspectives** 3. Ideas for future research projects 4. Vision and place for InterConnect approach
16:30	**Discussion and involvement**

Plan

- This talk:
  - Illustrate scientific opportunity
    - Enabling increased understanding differences in risk of diabetes and obesity between populations
  - Introduction to InterConnect approach
- This symposium:
  - Show InterConnect approach works
    - Set up takes some work but doable and producing scientifically interesting results of public health relevance
    - A foundation has been created – now build further

InterConnect

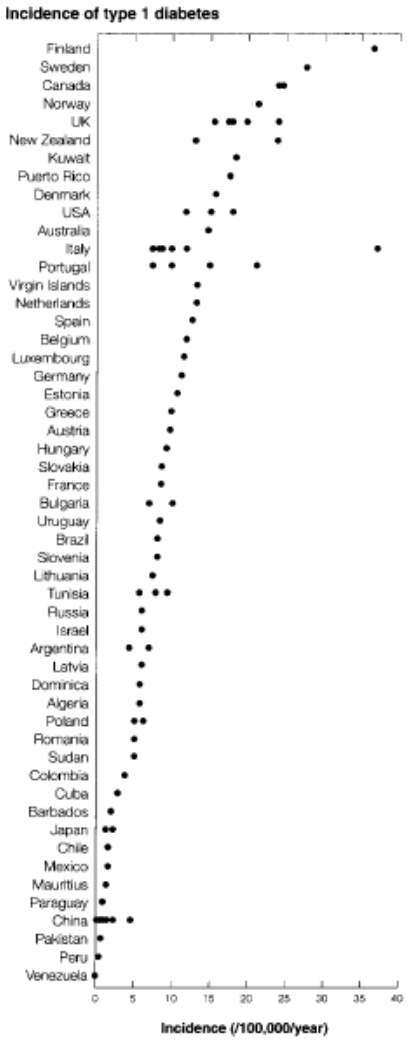
Global data for diabetes and obesity research

The InterConnect Project

Nick Wareham
InterConnect Co-ordinator, University of Cambridge, UK

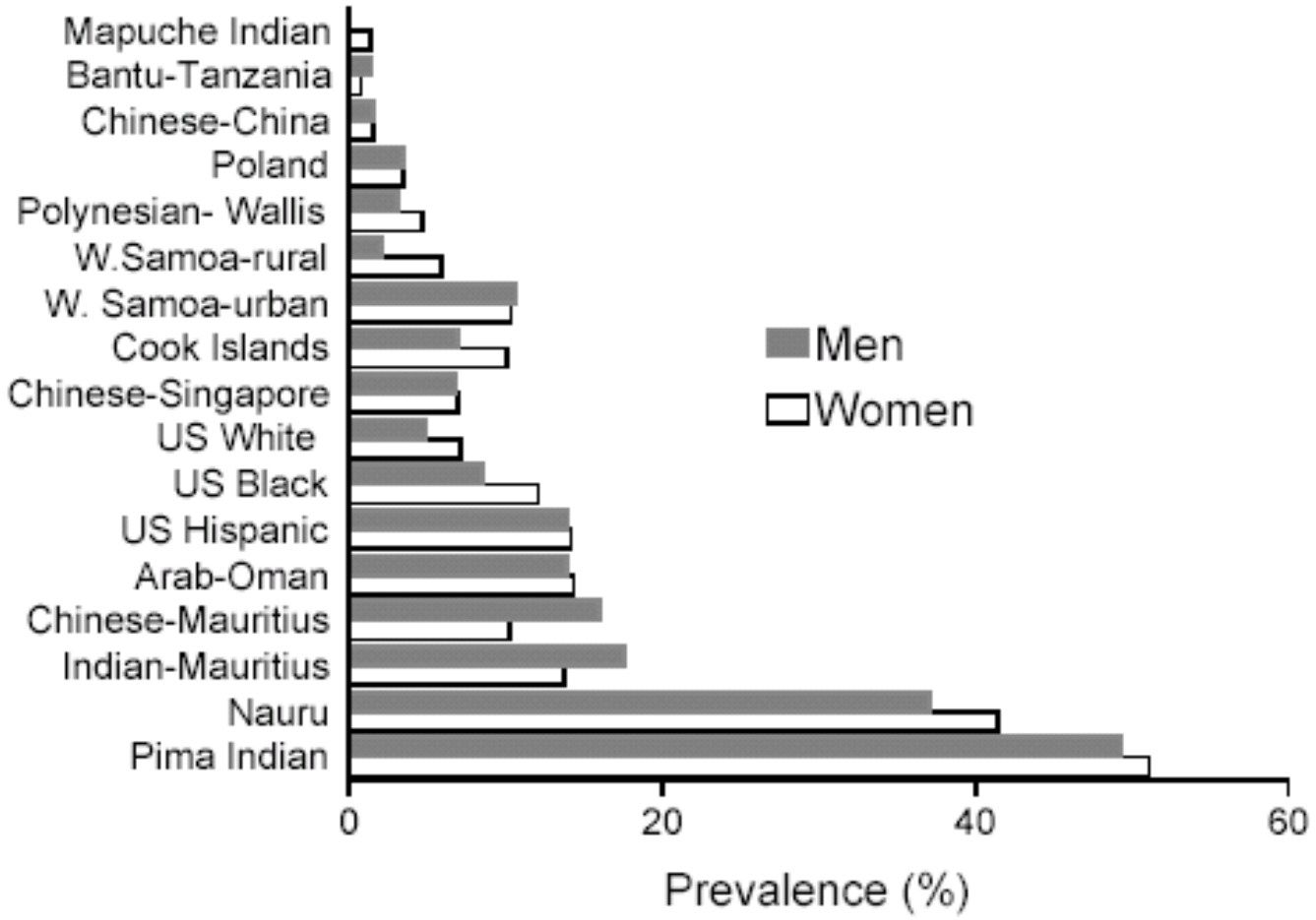
This project is funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 602068.

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



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

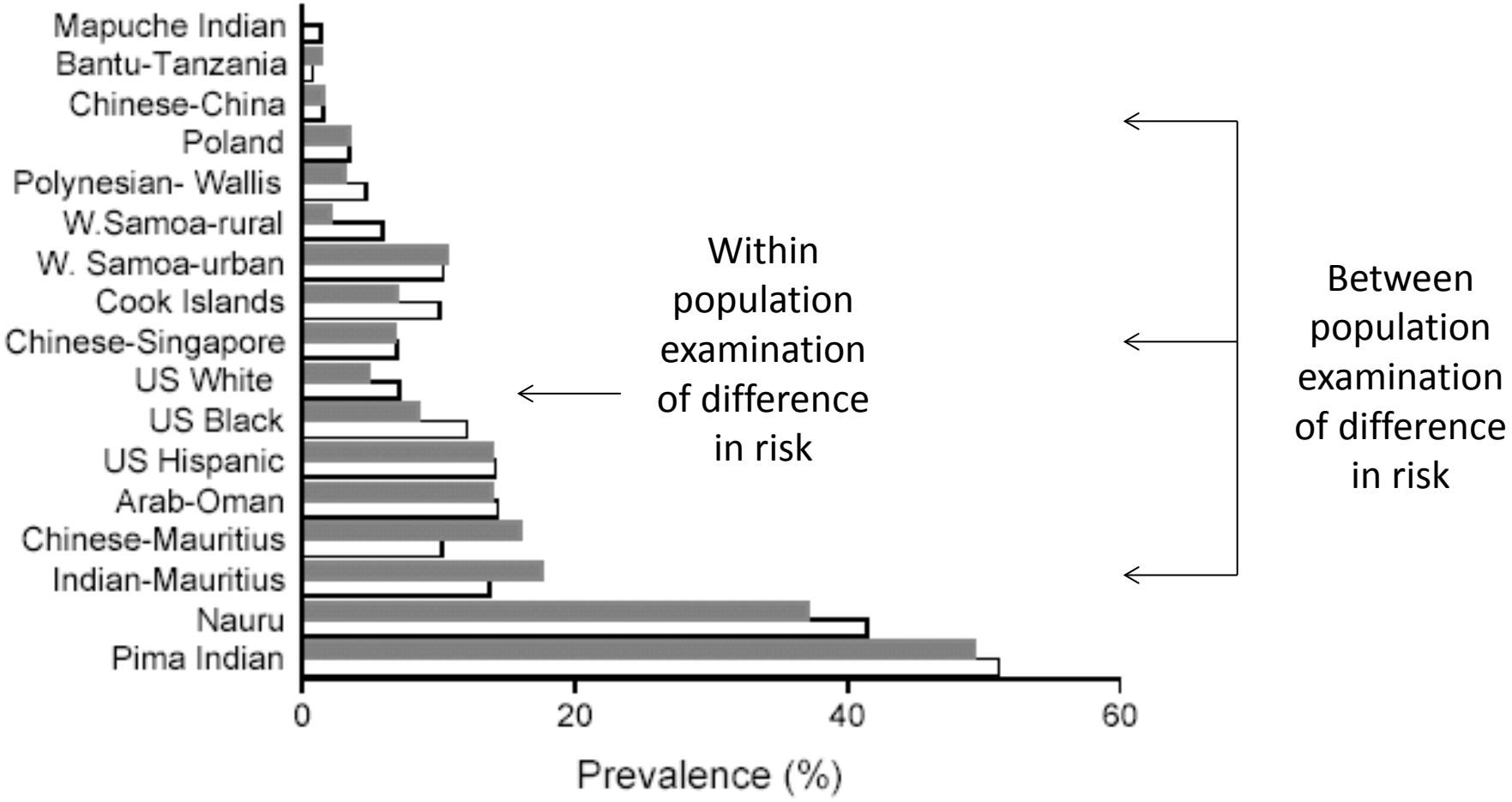
Design and cohort description of the InterAct Project: an examination of the interaction of genetic and lifestyle factors on the incidence of type 2 diabetes in the EPIC Study

The InterAct Consortium

Research groups in 8 countries; 26 centres

Source: Langenberg C et al, Diabetologia 2011

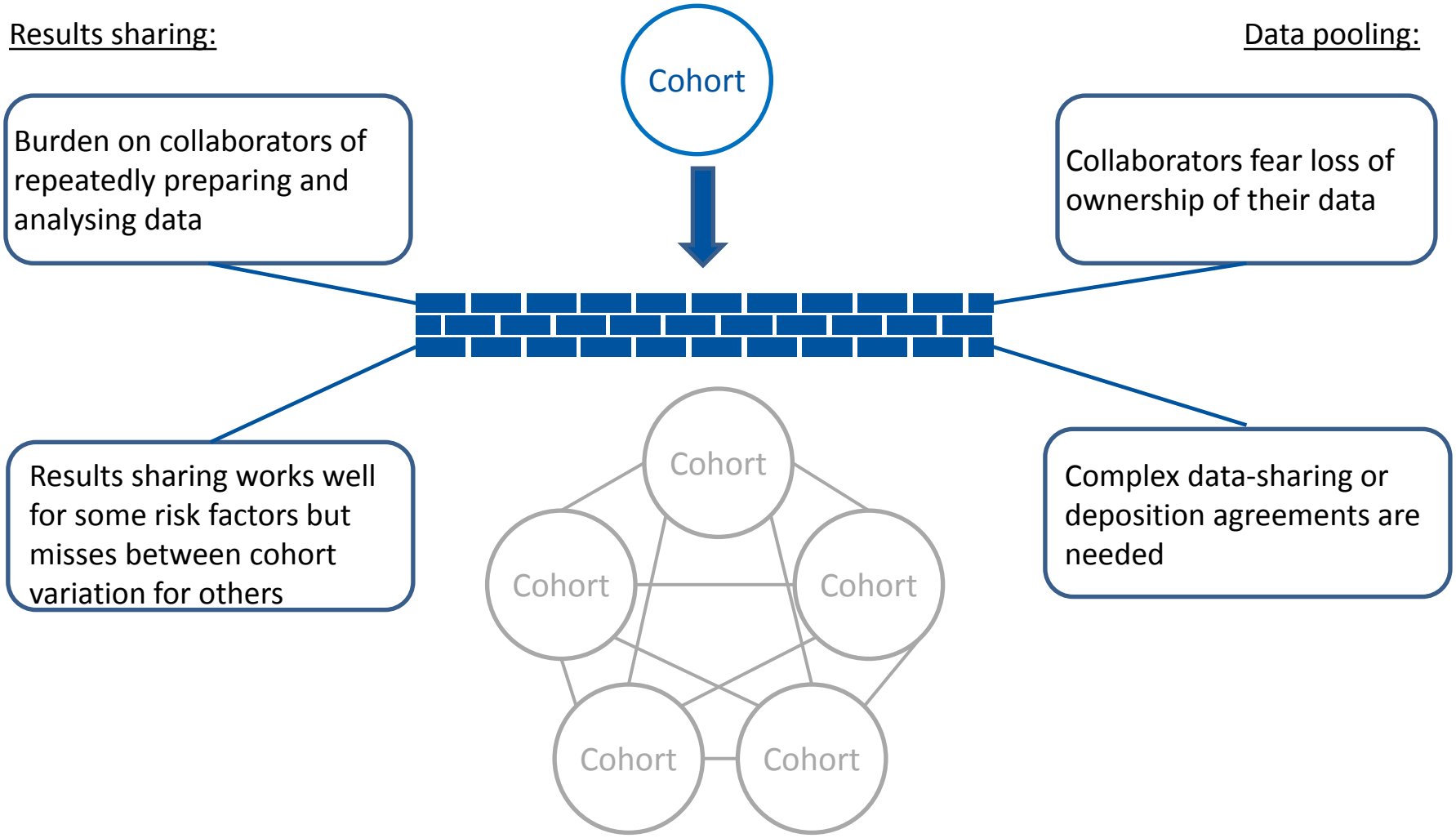
Phase 3: Moving from within-population investigation to the study of between-population differences



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

Barriers to cross-cohort analyses



InterConnect

- Goal to optimise use of existing data to enable cross-cohort analyses
  - Individual participant meta-analysis of pooled data from separate cohorts is analytically desirable
  - InterConnect aims to enable a solution without physical pooling of data by TAKING THE ANALYSIS TO THE DATA

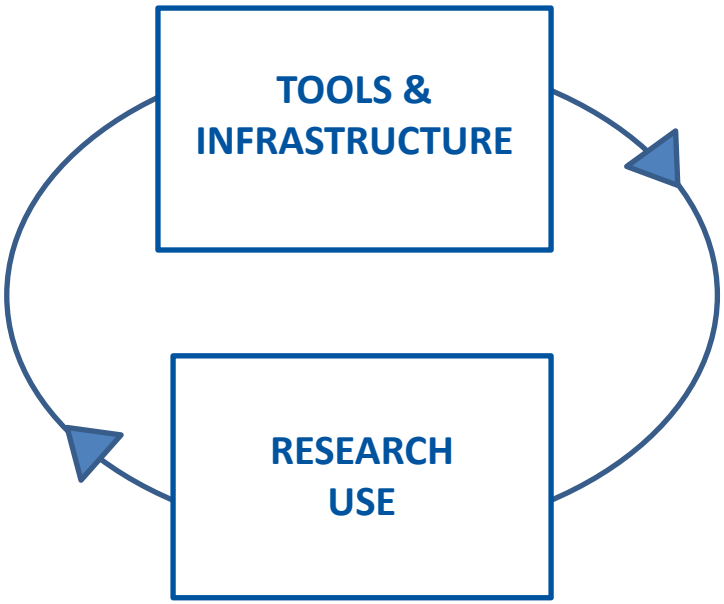
InterConnect is different…

- Goal is to enable others
- Creating an approach to optimise the use of existing data for cross-cohort analyses
  - currently constrained by limitations of conventional data sharing and approaches to meta-analysis

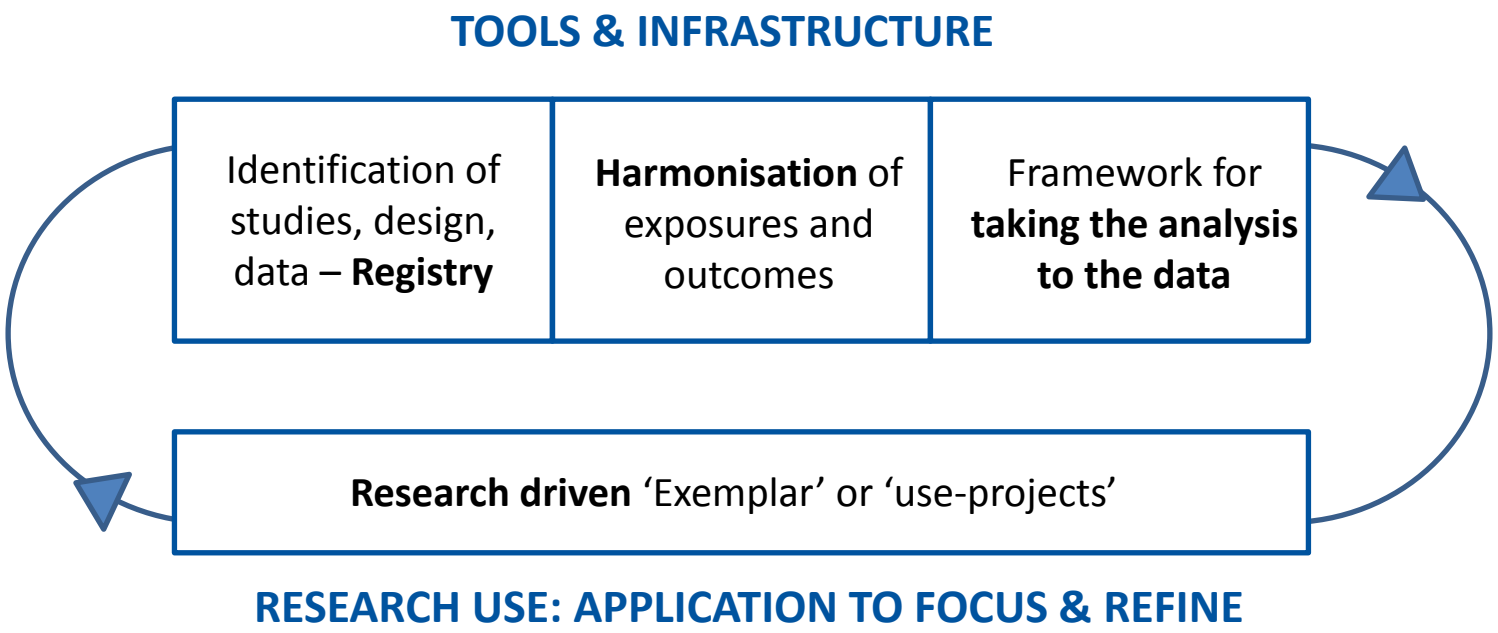
Creating change requires many actors

- Researchers – to see need, think useful, demonstrate value
- Stakeholders who are users of research evidence – create pull
- Funders – infrastructure, incentives for re-use of data

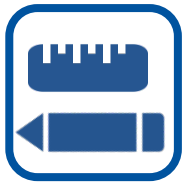
InterConnect: A bridging function



InterConnect: A bridging function



Identification of studies, design, data – **Registry**	**Harmonisation** of exposures and outcomes	Framework for **taking the analysis to the data**

	A catalogue of studies relating to diabetes and obesity
	Populations recruited to the study
	Biological samples stored or analysed
	The study design that was employed

InterConnect: Live study registry

www.interconnect-diabetes.eu

Identification of studies, design, data – **Registry**	**Harmonisation** of exposures and outcomes	Framework for **taking the analysis to the data**

Exemplar question: Study A
In a typical week, how many glasses of red wine (6 ounces) do you drink per day?
[___] Number of drinks per day

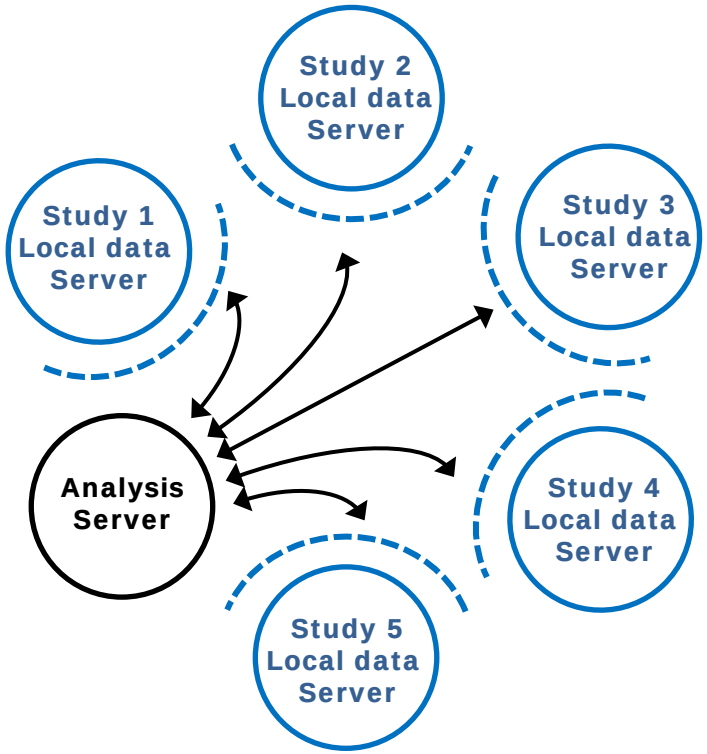
Exemplar question: Study B
In general, how many glasses of red wine do you drink per day over a week and weekend?
Week: [___] Number/day
Weekend: [___] Number/day

Exemplar question: Study C
In a typical week, how many glasses of red wine do you drink per day?
❑ 1–3
❑ 4–6
❑ 7–9
❑ 10 or more

→ Align to give a single exposure where possible

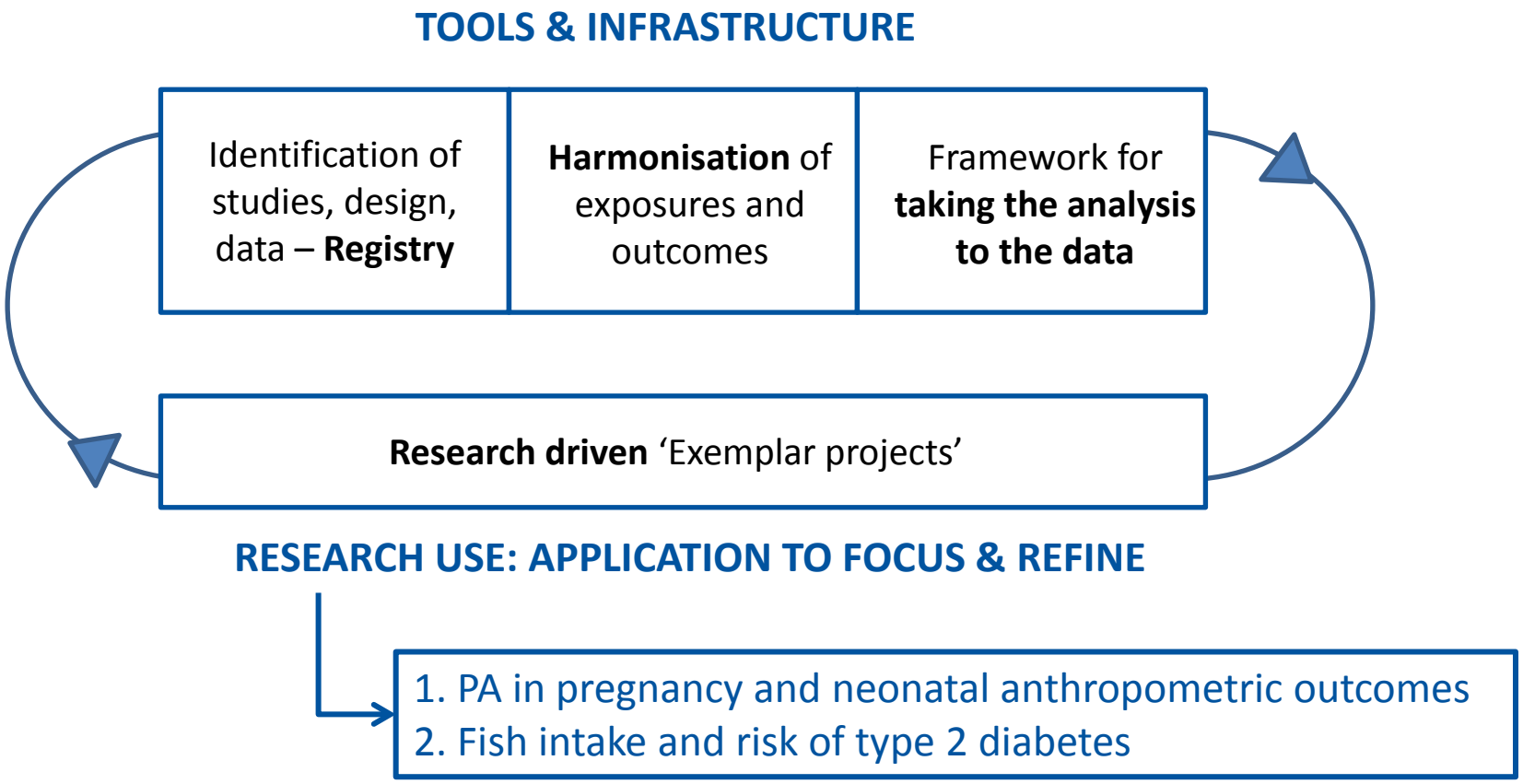
InterConnect software captures how the alignment is made so that it is both explicit and re-usable

Identification of studies, design, data – **Registry**	**Harmonisation** of exposures and outcomes	Framework for **taking the analysis to the data**



- Take the analysis to the data - federated analysis
- Data stay within the governance structure of the cohort
- Analytical instructions and non-identifying summary parameters allowed to pass between computers
- Any user with appropriate log in credentials can remotely access the analysis server to run analysis code

InterConnect: A bridging function



Programme

14:30	**The InterConnect Project**	
14:45	**Applying the InterConnect approach for federated meta-analysis** 1. Physical activity in pregnancy & neonatal anthropometric outcomes - *Why this question is important* - *Why federated meta-analysis is required* - *Harmonisation & set up for federated meta-analysis* - *Analysis plan and results* 2. Fish intake and risk of type 2 diabetes	Gernot Desoye Silvia Pastorino Tom Bishop Ken Ong
16.00	**Future perspectives**	
16:30	**Discussion and involvement**	

InterConnect

Global data for diabetes and obesity research

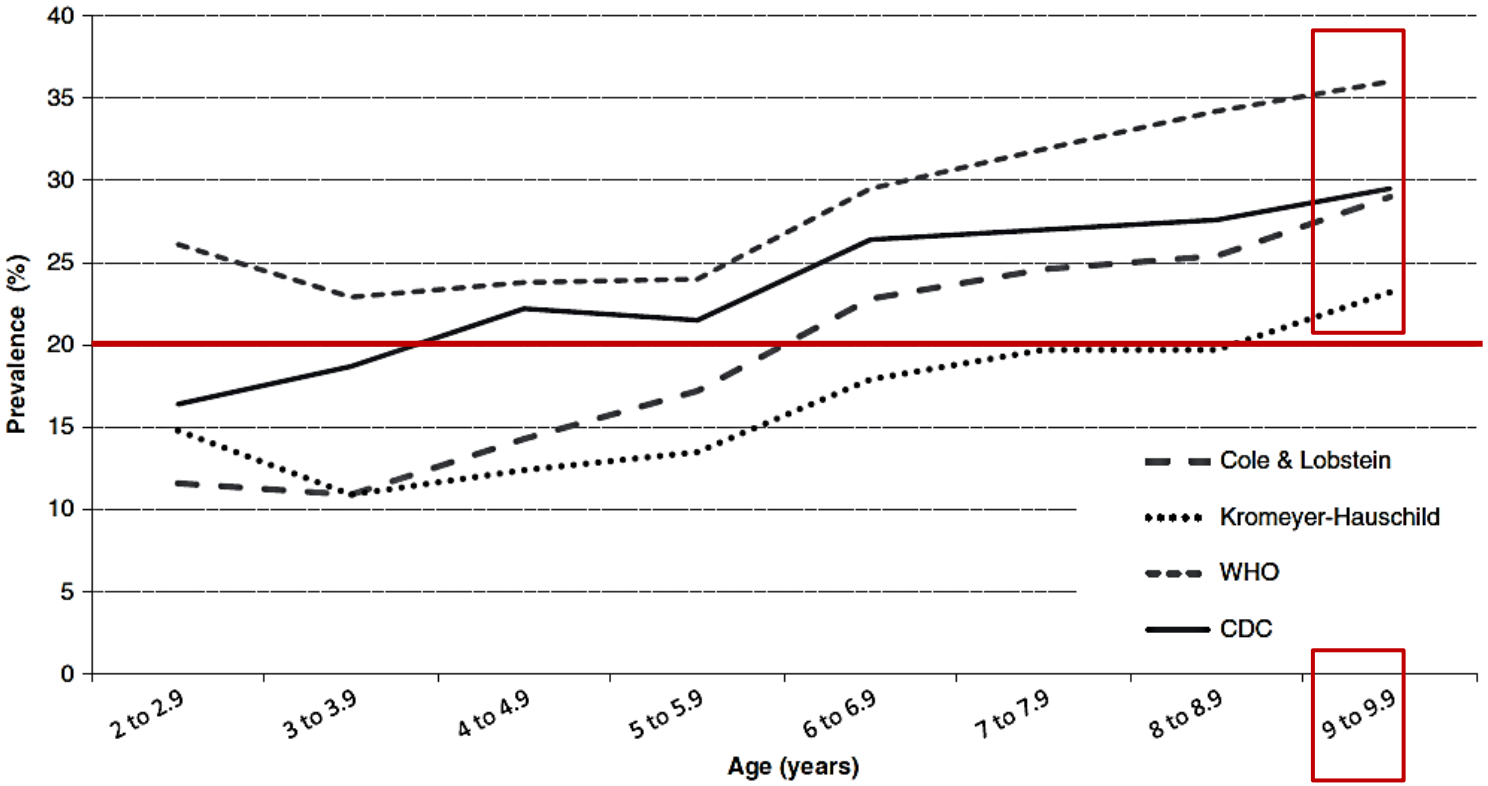
Effect of Physical Activity on Neonatal Anthropometric Outcomes:

Why this question is important

Gernot Desoye
Dept Obstetrics and Gynaecology, Medical University of Graz, Graz, Austria

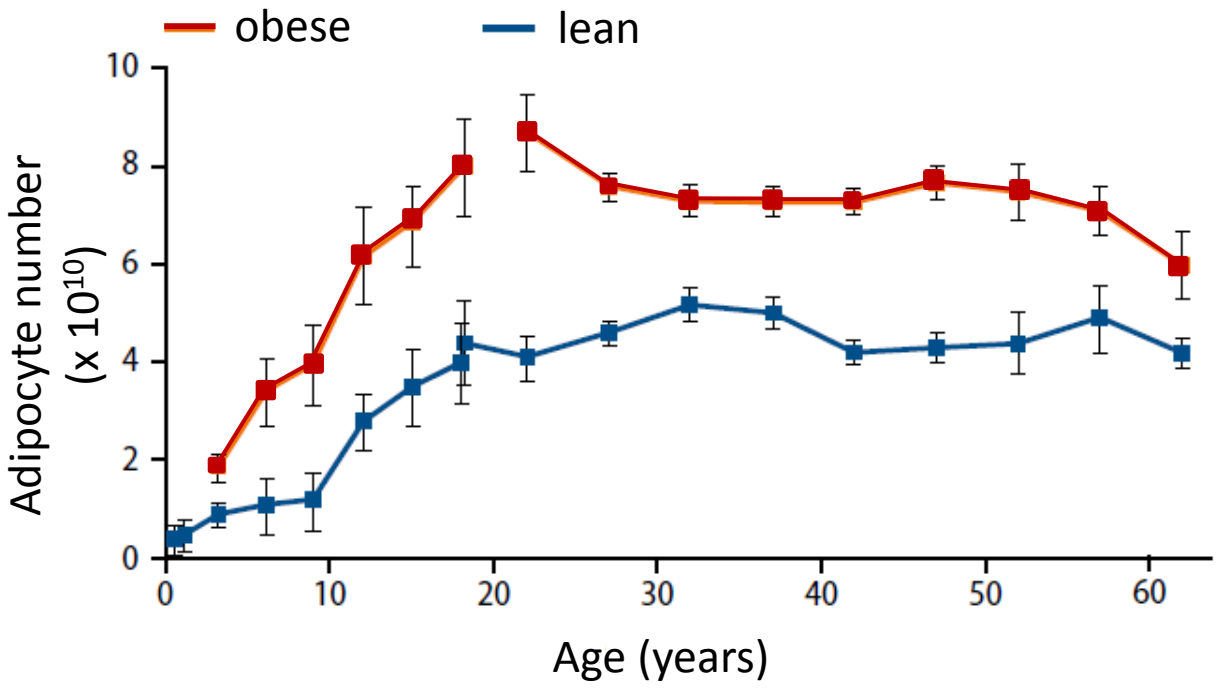
This project is funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 602068.

More than 20% of European children are Overweight/obese at age 10



Ahrens W et al (IDEFICS consortium), Int J Obes 38:S99-107, 2014

Total adipocyte cell number is established early in life and is greater in the obese



Spalding KL et al, Nature 453: 783-787, 2008

Higher risk of LGA vs AGA offspring for developing Metabolic Syndrome at 11 years of age (n=175)

Variable	Hazard Ratio	P Value	95% CI for Hazard Ratio
LGA vs AGA	**2.19**	.006	1.25-3.82

Metabolic Syndrome is defined as presence of ≥2 of 4 major components (obesity, hypertension, high TG or low HDL levels, glucose intolerance) ; Cox regression analysis

Boney CM et al, Pediatrics 115:e290-e296, 2005

Offspring born macrosomic have a higher risk for overweight/obesity at 7 years of age

BMI	Macrosomia ≥4,000g	Control	OR (95% CI)	P
Normal	63%	72 %	-	-
Overweight	22 %	17 %	**1.52** (1.24-1.86)	0.001
Obesity	14 %	11 %	**1.50** (1.19-1.92)	< 0.001

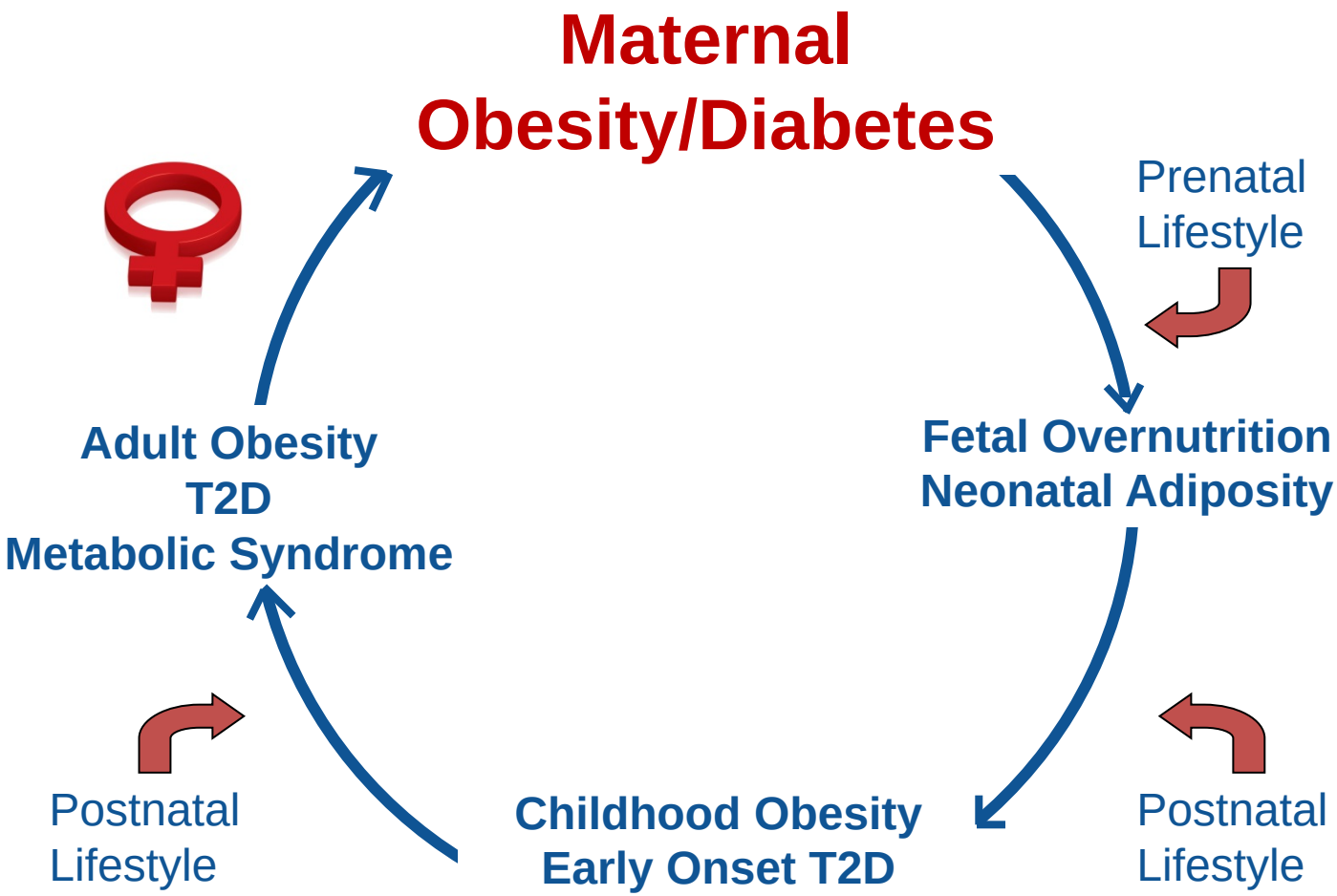
Gu S et al, J Biomed Res 26:235-240, 2012

Diabetes/glycosuria are risk factors for macrosomia (ALSPAC)

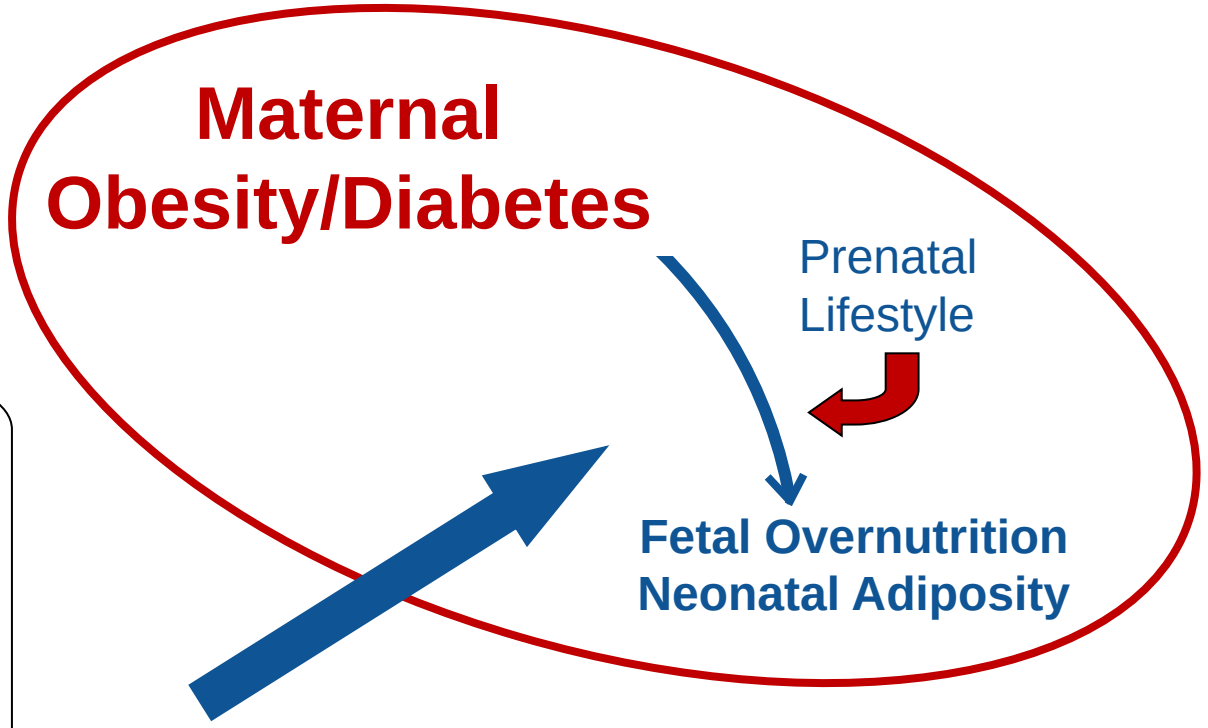
Maternal risk factor	Odds ratio for macrosomia (95% CI)
Existing diabetes	**3.56** (1.53-8.28)
GDM	**5.50** (1.18-10.30)
Glycosuria	**1.58** (1.18-2.12)

Lawlor DA et al, Diabetologia 53:89-97, 2010

Vicious cycle of diabesity



Vicious cycle of diabesity



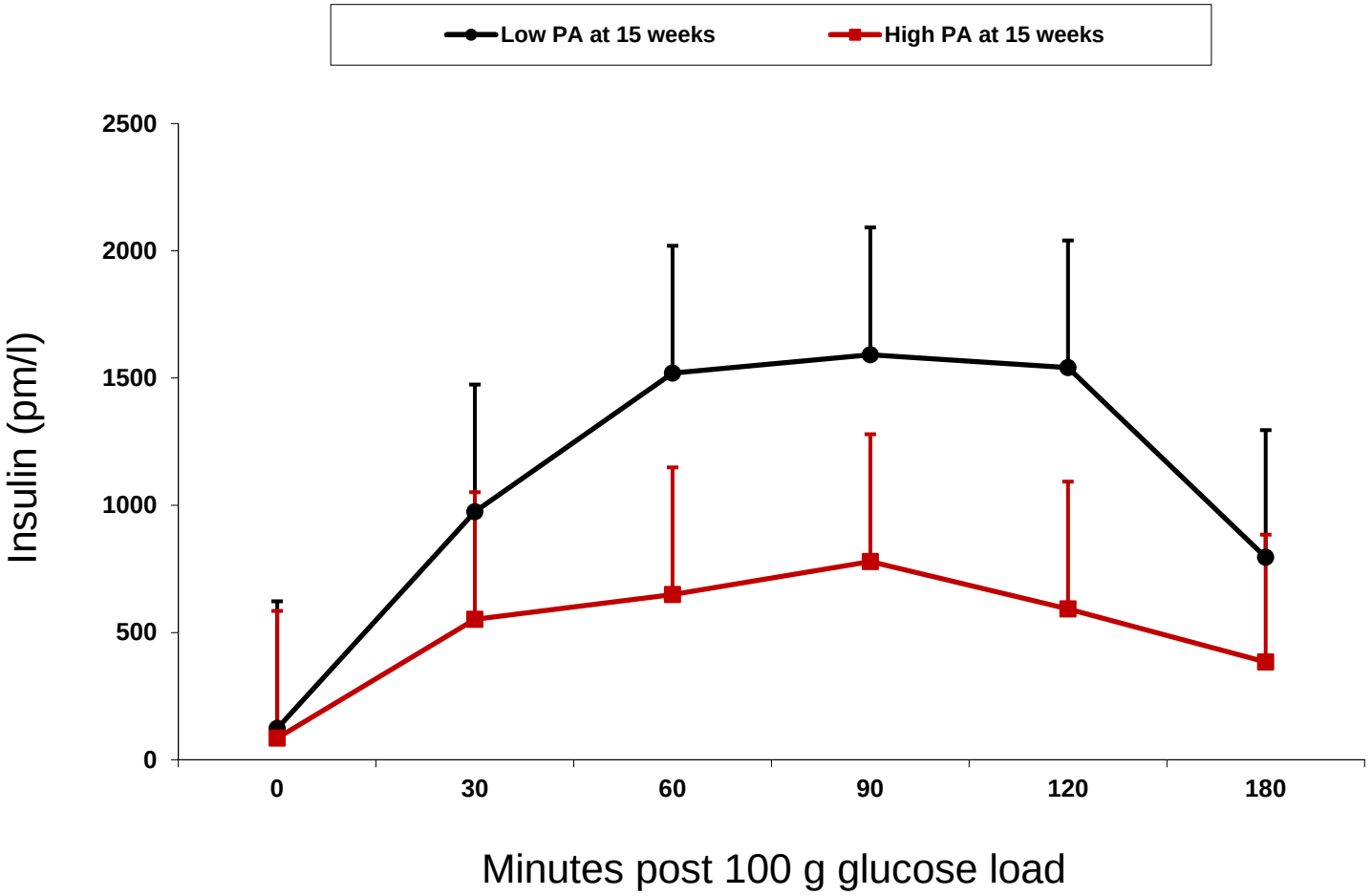
Environment

- Nutrition
- Physical Activity
- Inflammation
- Metabolism
- Endocrine Status
- Stress
- Infection

Interaction with

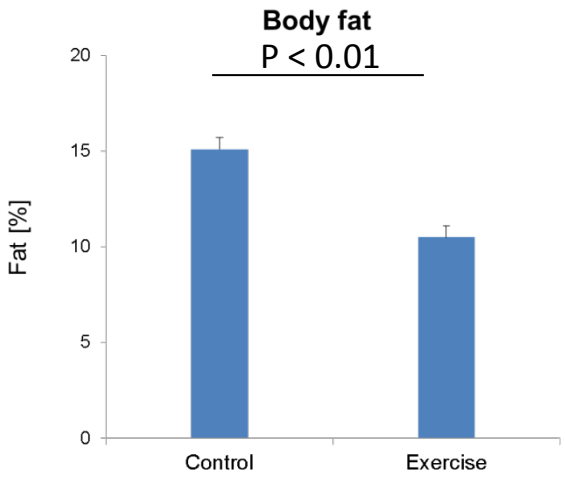
Fetal (Epi)Genome

PA at 15 wks improves insulin response (oGTT) at 32 wks in Overweight/obese women

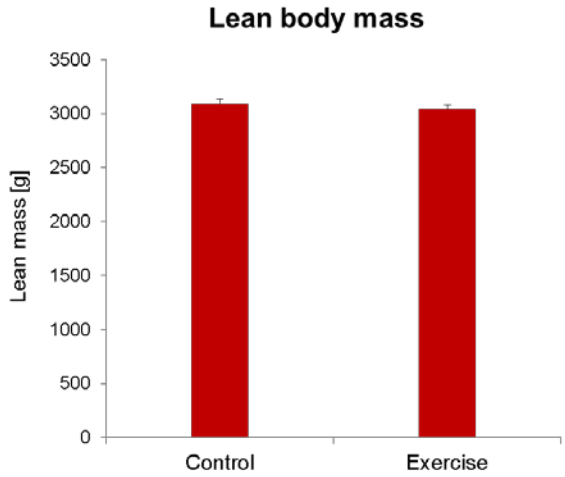


Van Poppel M et al, JCEM 98:2929-35, 2013

PA in pregnancy has long term beneficial effects on offspring

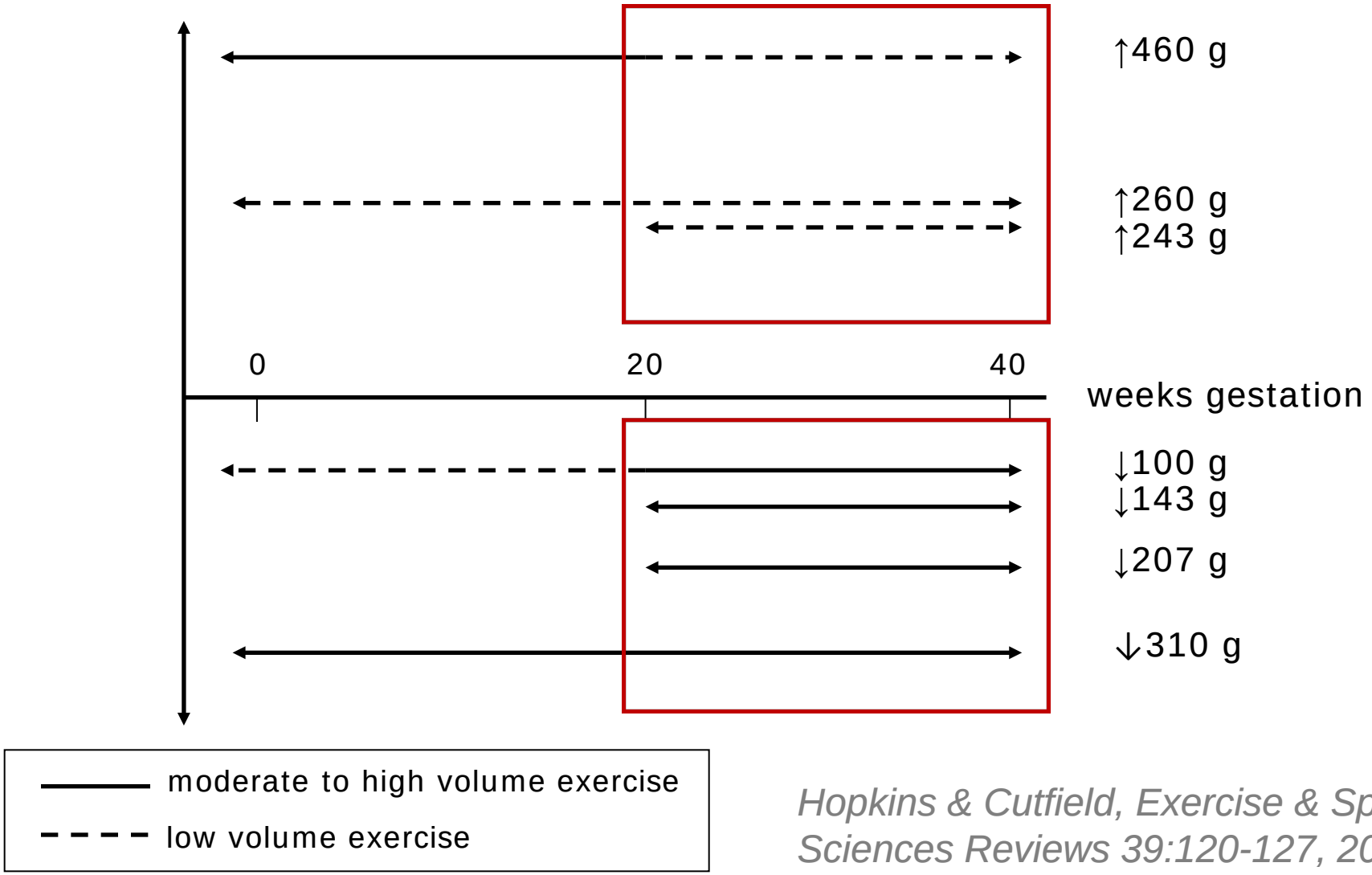


Birth



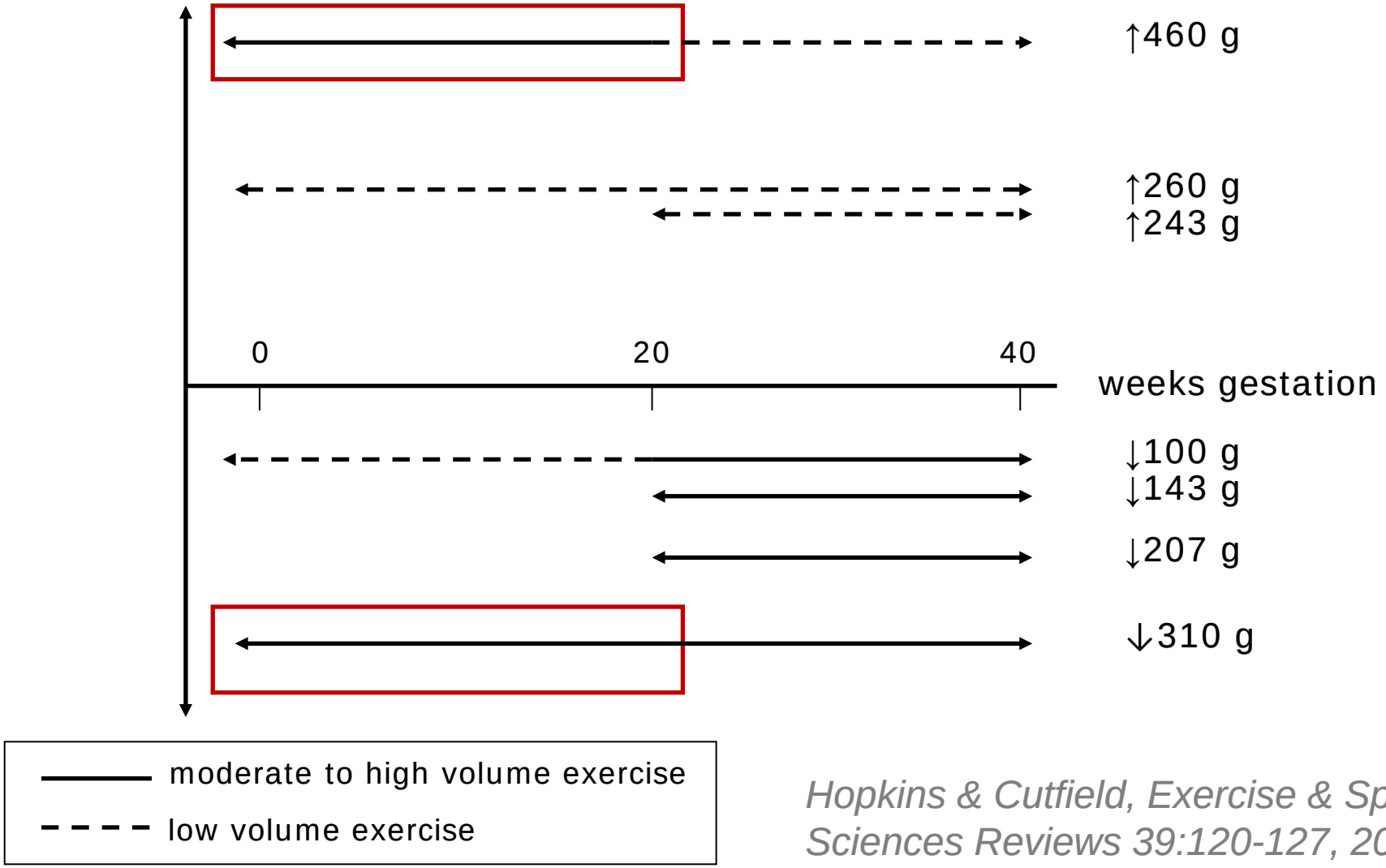
Clapp JF 3rd, J Pediatr 129:856-63, 1996

PA effect on birth weight depends on intensity and period in pregnancy



Hopkins & Cutfield, Exercise & Sport Sciences Reviews 39:120-127, 2011

PA effect on birth weight depends on intensity and period in pregnancy



Hopkins & Cutfield, Exercise & Sport Sciences Reviews 39:120-127, 2011

Summary

- Neonates born heavy (LGA/macrosomia) have a higher risk for being overweight/obese and to show early features of the metabolic syndrome in childhood
- Diabetes in pregnancy is a risk factor for heavy neonates
- Physical activity in pregnancy may improve maternal glucose tolerance
- Its effects on birth weight depend on exercise intensity/volume and the period in pregnancy, when mothers are physically active
- These results have so far been obtained in small ‘cohorts’ only

InterConnect

Global data for diabetes and obesity research

Systematic reviews of maternal physical activity in pregnancy and offspring birth size

Silvia Pastorino
MRC Epidemiology Unit

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

What is known?

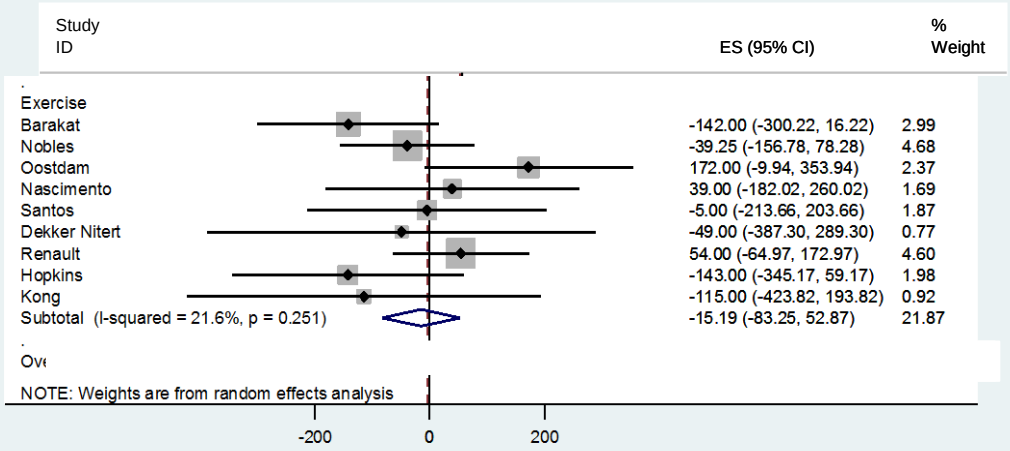
- Various studies investigated PA in pregnancy and offspring birth size, with conflicting and inconclusive results
- **Systematic reviews**:
  - Randomized controlled trials (RCTs):
    - *Wiebe et al, 2015*
    - *Sanabria-Martínez et al, 2015*
    - *Lisa Kafer (unpublished)*
  - Observational studies
    - *Schlussel, 2008*
    - *Bisson, 2016*

Systematic reviews of RCTs

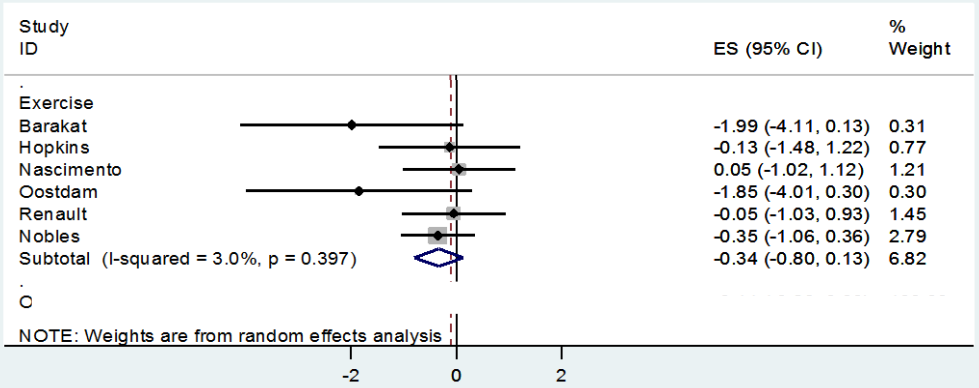
- *Two recent meta-analyses of maternal PA interventions suggest modest decreases in birth weight and risk of LGA
- **However a recent systematic review of interventions among overweight and obese women found no effect
- High heterogeneity in effect sizes
- Effects of Volume / Intensity of PA not investigated

**Wiebe et al, 2015; Sanabria-Martínez et al, 2015*
*** Lisa Kafer (unpublished)*

Effect of interventions among overweight and obese women on BW (g) and LGA



Difference in Birth weight by Intervention



ln(OR) for LGA by Intervention

Lisa Kafer (unpublished)

Systematic reviews of Observational Studies (1) Schlussel et al, 2008

	Birth weight (BW)	Large for gestational age (LGA)
Negative association	3	1
No association	5	
Positive association	2	
Negative at high levels, Positive at moderate levels	1	

- Studies lacked standardization as to the type of activities evaluated; not possible to pool results by meta-analysis

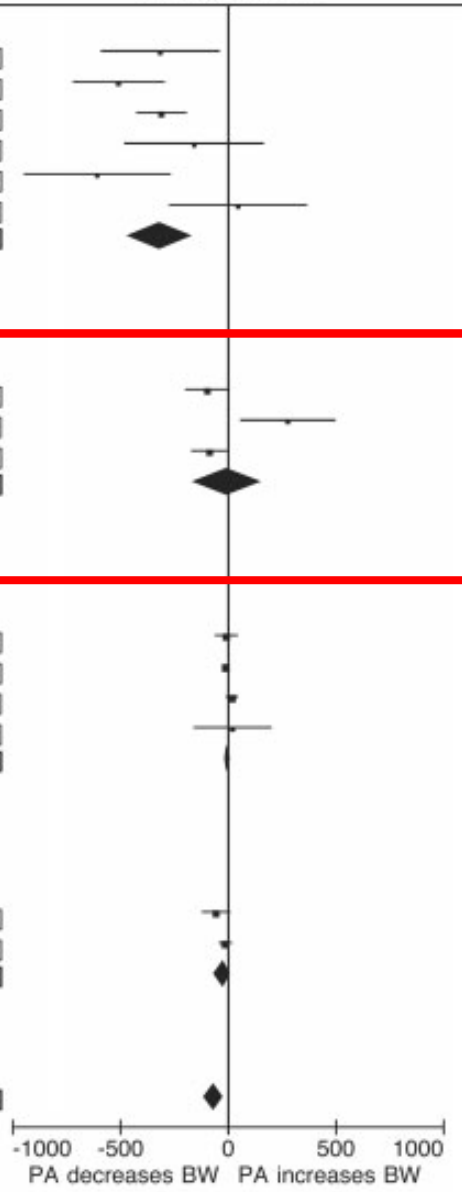
Systematic reviews of Observational Studies (2) Bisson et al, 2016

	Birth weight (BW)	LGA or Macrosomia	% Body Fat
Negative association	8	8	2
No association	25	5	
Positive association	4		

- Most studies found no association with BW
- LTPA associated with lower OR of LGA or Macrosomia, and lower % Body Fat
- Notably, 19 of 42 studies did not adjust for any confounder

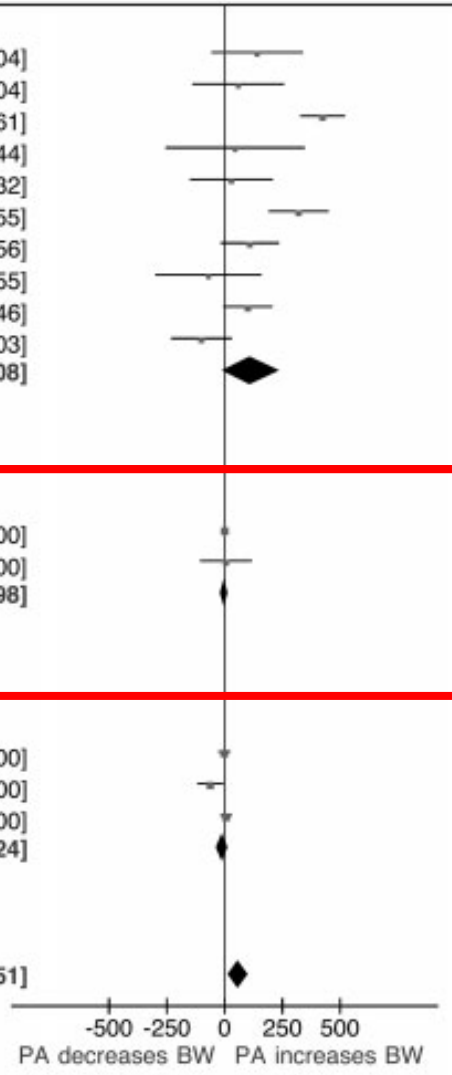
Association between pregnancy PA and offspring BW – High PA levels

Study or Subgroup	Mean Difference [g]	SE	High level Total	Low level Total	Weight	Mean Difference Random, 95% CI
2.1 High volume vs low volume - crude						
Bell 1995	-315	141	58	41	2.2%	-315.00 [-591.35, -38.65]
Clapp 1984	-509	108.8	29	152	3.3%	-509.00 [-722.35, -295.65]
Clapp and Capeless 1990	-310	59.5	77	55	7.0%	-310.00 [-426.62, -193.38]
Duncombe 2006	-158.6	165.1	27	17	1.7%	-158.60 [-482.38, 165.18]
Perkins 2007	-608	174.4	12	13	1.5%	-608.00 [-949.82, -266.18]
Rice 1991	45.4	163.5	12	11	1.7%	45.40 [-275.05, 365.85]
Subtotal (95% CI)			**215**	**289**	**17.4%**	**-319.33 [-472.21, -166.45]**
Heterogeneity: Tau² = 19373.36; Chi² = 11.71, df = 5 (P = 0.04); I² = 57%						
Test for overall effect: Z = 4.09 (P < 0.0001)						
2.2 High volume vs low volume - adjusted						
Harrod 2014 [a]	-97	52.9	206	206	7.7%	-97.00 [-200.68, 6.68]
Hatch 1993	276	113	15	185	3.1%	276.00 [54.52, 497.48]
Magann 2001	-86.5	43.7	238	217	8.9%	-86.50 [-172.15, -0.85]
Subtotal (95% CI)			**459**	**608**	**19.7%**	**-5.00 [-161.86, 151.86]**
Heterogeneity: Tau² = 14416.18; Chi² = 9.70, df = 2 (P = 0.008); I² = 79%						
Test for overall effect: Z = 0.06 (P = 0.95)						
2.3 High duration vs low duration						
Hegaard 2010	-9	27.0	289	3672	11.1%	-9.00 [-62.00, 44.00]
Juhl 2010	-11	9.4	2236	49929	12.8%	-11.00 [-29.42, 7.42]
Nieuwenjuijsen 2002 [b]	16.7	14.4	1131	6744	12.5%	16.70 [-11.52, 44.92]
Sternfeld 1995	20	92.6	25	122	4.2%	20.00 [-161.49, 201.49]
Subtotal (95% CI)			**3681**	**60467**	**40.5%**	**-3.06 [-17.82, 11.70]**
Heterogeneity: Tau² = 0.00; Chi² = 2.71, df = 3 (P = 0.44); I² = 0%						
Test for overall effect: Z = 0.41 (P = 0.68)						
2.4 High intensity vs low-moderate intensity						
Jukic 2012	-57	34.1	224	367	10.1%	-57.00 [-124.00, 10.00]
Rose 1991 [c]	-13	16.3	1281	17927	12.3%	-13.00 [-44.95, 18.95]
Subtotal (95% CI)			**1505**	**18294**	**22.4%**	**-24.74 [-62.88, 13.40]**
Heterogeneity: Tau² = 250.87; Chi² = 1.35, df = 1 (P = 0.25); I² = 26%						
Test for overall effect: Z = 1.27 (P = 0.20)						
Total (95% CI)			**5860**	**79658**	**100.0%**	**-69.85 [-114.75, -24.96]**
Heterogeneity: Tau² = 4006.44; Chi² = 80.59, df = 14 (P < 0.00001); I² = 83%						
Test for overall effect: Z = 3.05 (P = 0.002)						
Test for subgroup differences: Chi² = 17.10, df = 3 (P = 0.0007), I² = 82.5%						



Association between pregnancy PA and offspring BW – Moderate PA levels

Study or Subgroup	Mean Difference [g]	SE	Moderate level Total	Low level Total	Weight	Mean Difference Random, 95% CI
3.1 Moderate volume vs low volume - crude						
Botkin 1991	140.5	101.3	19	25	3.6%	140.50 [-58.04, 339.04]
Clapp 1984	59	101.5	47	152	3.6%	59.00 [-140.04, 258.04]
Downs 2007	424.4	49.6	41	18	7.6%	424.40 [327.19, 521.61]
Duncombe 2006	45.6	153.4	24	17	1.9%	45.60 [-255.24, 346.44]
Horns 1996	29	92	48	53	4.1%	29.00 [-151.32, 209.32]
Jahromi 2011	320	66.1	70	62	6.0%	320.00 [190.45, 449.55]
Johnson 1994	109.2	64.9	139	95	6.1%	109.20 [-18.16, 236.56]
Melzer 2010	-70	117.6	27	17	2.9%	-70.00 [-300.55, 160.55]
Morgan 2014	100	54.8	126	144	7.1%	100.00 [-7.46, 207.46]
Portela 2014 [a]	-100	67.3	37	19	5.9%	-100.00 [-232.03, 32.03]
Subtotal (95% CI)			578	602	48.7%	115.30 [-6.48, 237.08]
Heterogeneity: Tau² = 30923.55; Chi² = 58.39, df = 9 (P < 0.00001); I² = 85%						
Test for overall effect: Z = 1.86 (P = 0.06)						
3.2 Moderate volume vs low volume - adjusted						
Juhl 2010	0	7.1	4447	49929	11.7%	0.00 [-14.00, 14.00]
Sternfeld 1995	6	57.6	95	122	6.8%	6.00 [-107.00, 119.00]
Subtotal (95% CI)			4542	50051	18.5%	0.09 [-13.80, 13.98]
Heterogeneity: Tau² = 0.00; Chi² = 0.01, df = 1 (P = 0.92); I² = 0%						
Test for overall effect: Z = 0.01 (P = 0.99)						
3.3 Moderate intensity vs low intensity						
Hegaard 2010 [b]	-1	13.2	2384	1998	11.4%	-1.00 [-27.00, 25.00]
Jukic 2012	-62	28.5	527	367	10.0%	-62.00 [-118.00, -6.00]
Rose 1991	7	12.7	17927	2134	11.4%	7.00 [-18.00, 32.00]
Subtotal (95% CI)			20838	4499	32.8%	-9.13 [-38.50, 20.24]
Heterogeneity: Tau² = 383.21; Chi² = 4.90, df = 2 (P = 0.09); I² = 59%						
Test for overall effect: Z = 0.61 (P = 0.54)						
Total (95% CI)			25958	55152	100.0%	61.45 [16.40, 106.51]
Heterogeneity: Tau² = 4461.44; Chi² = 110.80, df = 14 (P < 0.00001); I² = 87%						
Test for overall effect: Z = 2.67 (P = 0.008)						
Test for subgroup differences: Chi² = 3.81, df = 2 (P = 0.15), I² = 47.5%						



Limitations of existing systematic reviews

- High heterogeneity due to:
  - Different consideration of confounding between studies; many studies were unadjusted
  - Different PA exposures:
    - Different domains: total PA, LTPA, occupational PA
    - Different volume or intensity
    - Categorisation not standardised
    - Different timings of PA during pregnancy
- Publication bias not tested

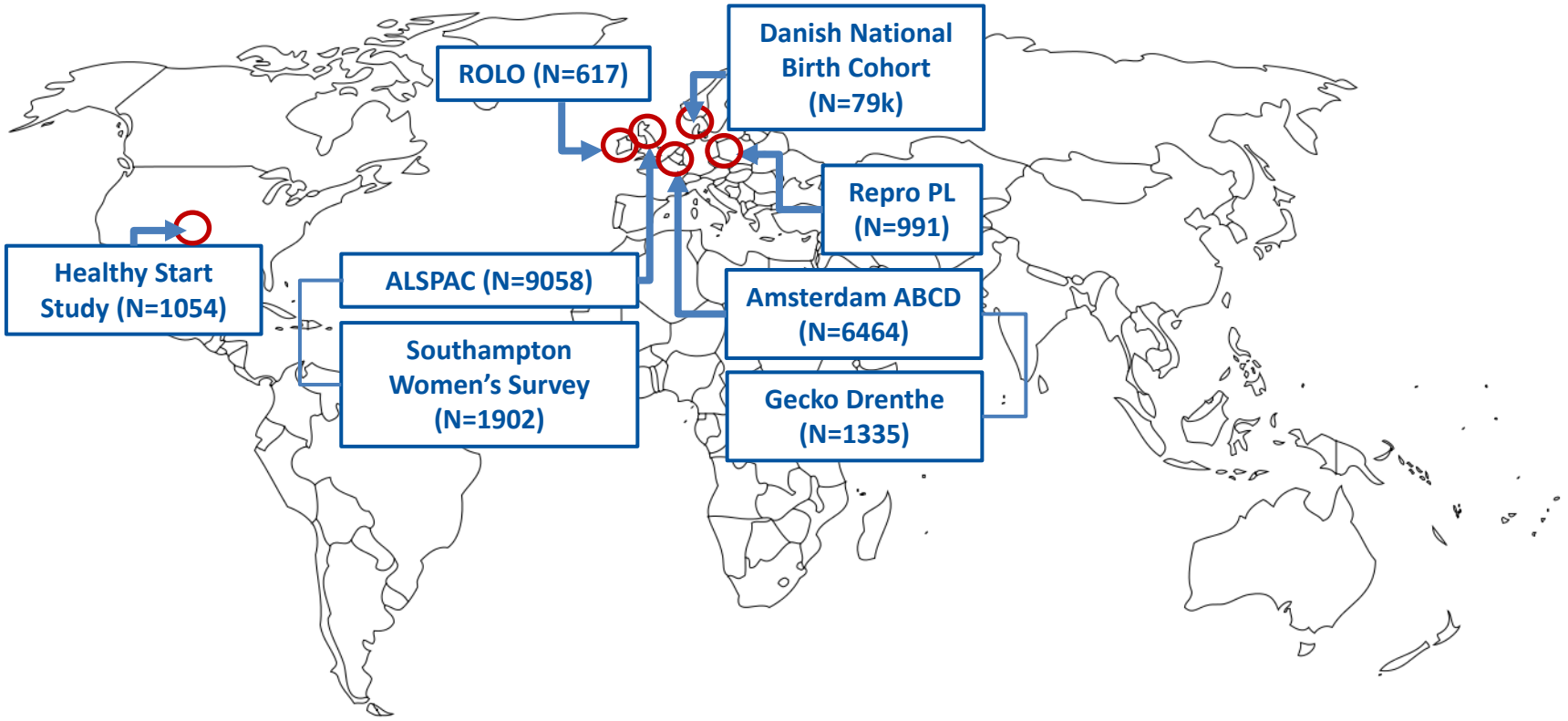
Alternative approaches

- Result sharing
  - Burden on collaborators of preparing and analysing data
  - More difficult to standardize measures across studies
- Data pooling
  - Collaborators fear loss of ownership of their data
  - Complex data-sharing agreements
- Federated meta-analyses
  - Data stay within the governance structure of the cohorts
  - Analytical instructions and non-identifying summary parameters allowed to pass between computers

Why using a federated meta-analysis

- Allows an Individual participant meta-analysis without physical pooling of data
- Reduces heterogeneity by:
  - Harmonising physical activity measures
  - Including the same number and types of confounders
- Allows investigation of:
  - Modifying factors
  - Different PA domains
  - Shape of the association and thresholds
  - Timing of PA in pregnancy (1^{st} or 3^{rd} trimester)
- Eliminate publication bias

Studies participating in the InterConnect physical activity in pregnancy project



InterConnect

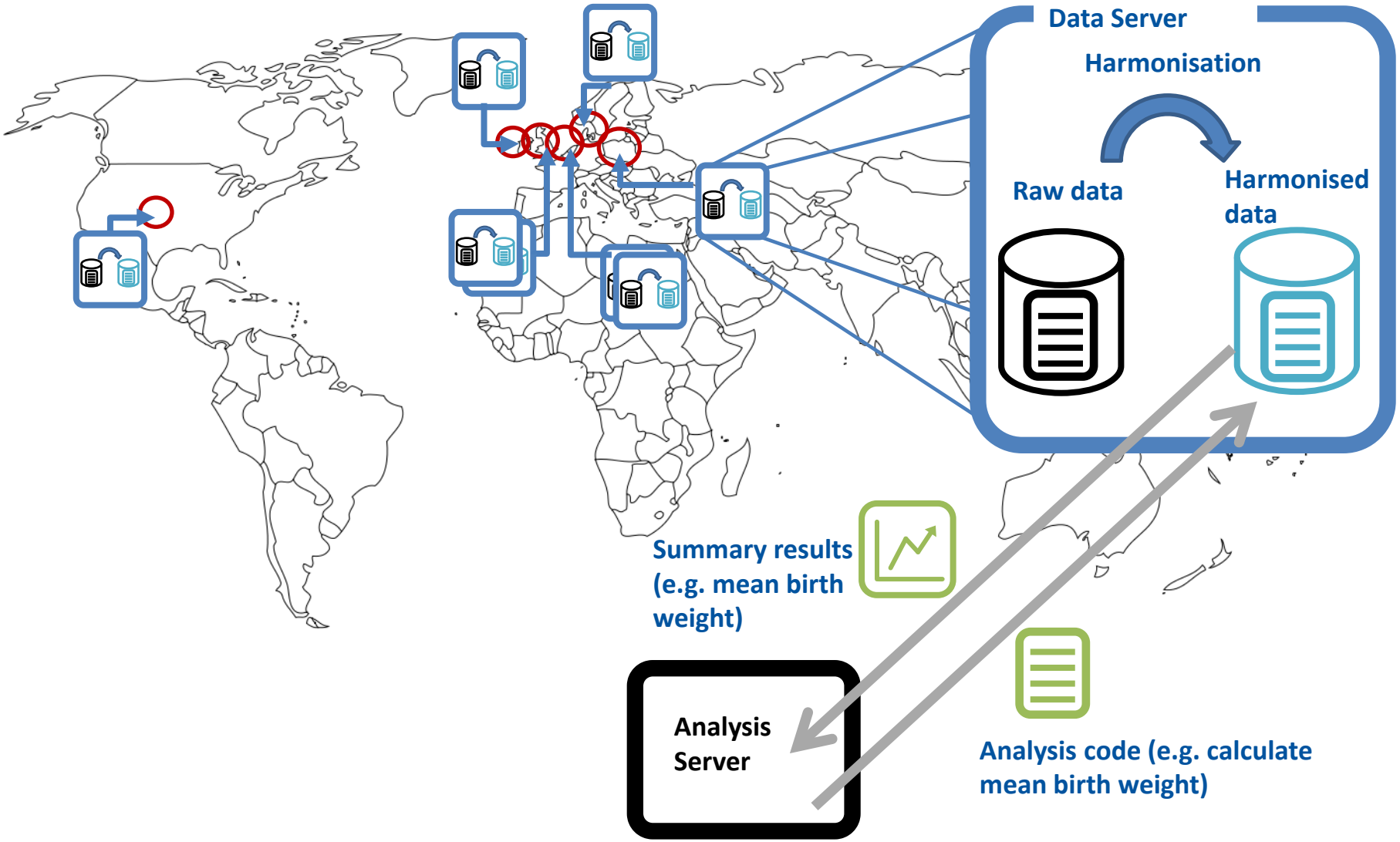
Global data for diabetes and obesity research

Technical set up & Harmonisation for federated meta-analysis

Tom Bishop,
Technical Lead, MRC Epidemiology Unit, University of Cambridge, UK

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

Seven participating studies have set up a server & harmonised data ready for analysis



Harmonisation example: Duration of moderate-vigorous leisure time physical activity

ALSPAC:

Question in questionnaire	**Q1. HOW MUCH do you do the following at present?** • Jogging • Aerobic • Antenatal exercises • Keep fit exercises • Yoga • Squash • Tennis/badminton • Swimming • Brisk walking • Weight training • Cycling • Other exercises
Frequency/duration	>7h/w, 2-6h/w, <1h/w, Never h/w = hours/week
Harmonised variable	**Duration of moderate/vigorous LTPA (h/w)**
Harmonisation rule	1. Convert: >7h/w to 7h/w 2-6h/w to 4h/w <1h/w to 0.5h/w Never to 0h/w 2. Sum up hours/w for all activities over 3 MET* (i.e. exclude antenatal exercises)
Missing data rule	Count single missing activity durations as 0 duration. If all activity durations missing, then mark participant as missing.

*MET = Metabolic Equivalent of Task

Harmonisation example: Duration of moderate-vigorous leisure time physical activity

ABCD:

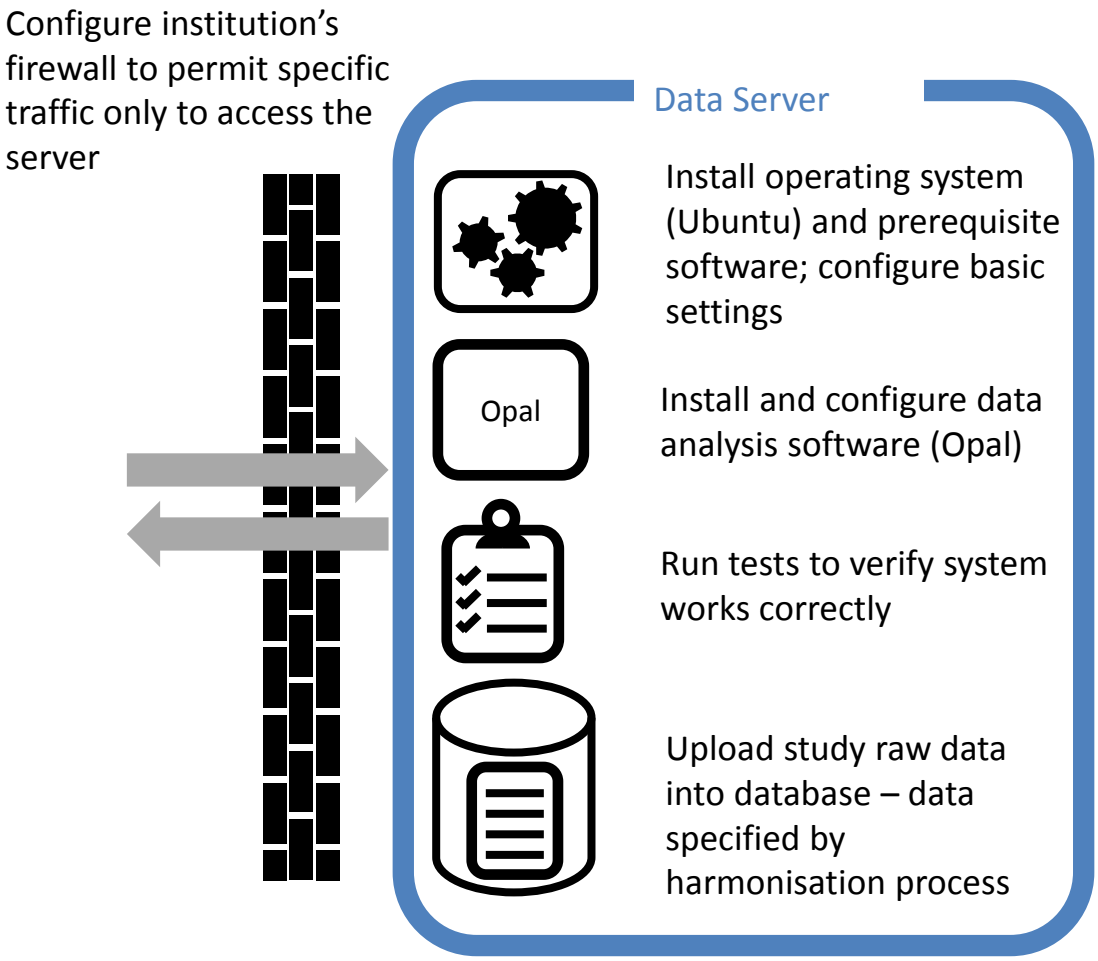
*MET = Metabolic Equivalent of Task

Question in questionnaire	**In your spare time did you:** **1.** Did you take **walks** for fun in the past week? **2.** Did you **ride a bicycle** in the past week? **3.** Did you **play sport**s in the past week? (for example: tennis, handball, gymnastics, fitness, skating, and swimming) **For each question: At what PACE do you usually do this?** • relaxed pace • average pace • brisk pace **For each question: FOR HOW LONG do you usually do this?**
Frequency/duration	Continuous value (mins/week)
Harmonised variable	**Duration of moderate/vigorous exercise (h/w)**
Harmonisation rule	1. Using Q1-3, sum up mins/w for all activities over 3 MET* (i.e. exclude relaxed walking) 2. Convert to hours (divide by 60)
Missing data rule	Count single missing activity durations as 0 duration. If pace is missing, assume relaxed. If all activity durations missing, then mark participant as missing.

The task and challenges of harmonisation are not unique to InterConnect

- Harmonisation was completed on all studies for all variables required for analysis
- Data pooling and results sharing for multiple studies also require this type of work
- InterConnect allows decisions & rules to be recorded for transparency & reuse

The InterConnect team supported the server set up & data upload for each study



Harmonisation rules were implemented in code on each study's server

Question in questionnaire	Q1. How much do you do the following at present? • Jogging • Aerobic • Antenatal exercises • Keep fit exercises • Yoga • Squash • Tennis/badminton • Swimming • Brisk walking • Weight training • Cycling • Other exercises
Frequency/duration	>7h/w, 2-6h/w, <1h/w, Never h/w = hours/week
Target variable	**Duration of moderate/vigorous LTPA (h/w)**
Harmonisation rule	1. Convert: >7h/w to 7h/w 2-6h/w to 4h/w <1h/w to 0.5h/w Never to 0h/w 2. Sum up hours/w for all activities over 3 MET (i.e. exclude antenatal exercises)
Missing data rule	Count single missing activity durations as 0 duration. If all activity durations missing, then mark participant as missing.

Harmonisation rules

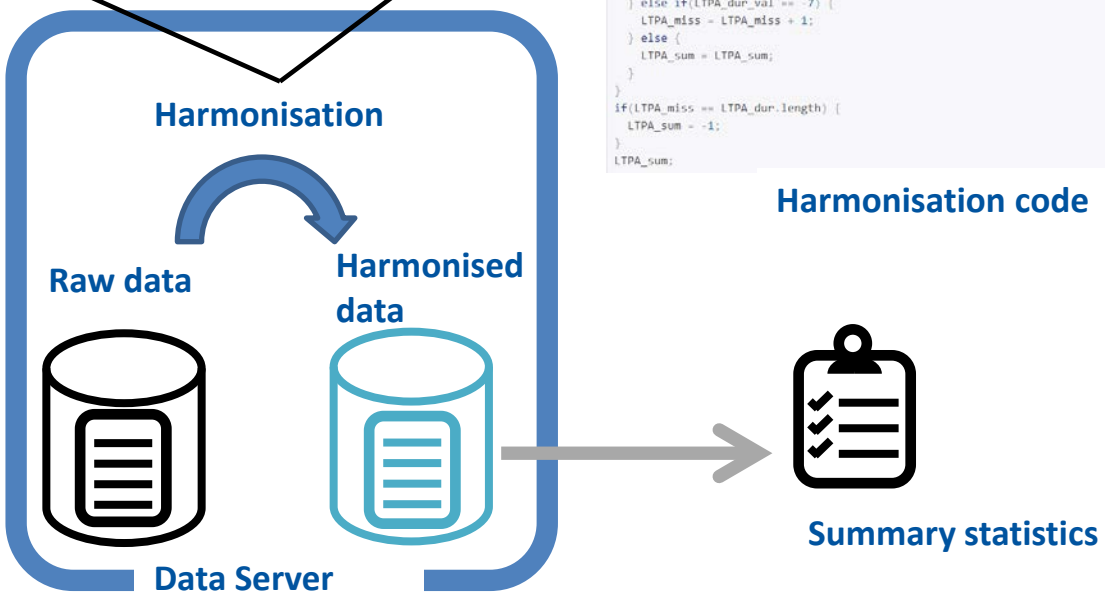
```
var LTPA_dur = ['b050', 'b051','b052','b053','b054','b055','b056','b057','b058','b059','b060','b061'];
var i;
var LTPA_dur_val = newValue(0, 'decimal');

var val_to_add = newValue(0, 'decimal');
var LTPA_sum = newValue(0, 'decimal');
var LTPA_miss = newValue(0, 'integer');

//no imputation - missing values counted as 0
// if they have a -7 for every answer, they didn't do the questionnaire so count as missing
// if they have -1s then they did the questionnaire but didnt fill anything in - assume no activity

for (i = 0; i < LTPA_dur.length; i++) {
  LTPA_dur_val = $(LTPA_dur[i]).map({
    '1': '7',
    '2': '4',
    '3': '0.5',
    '4': '0',
    '9': '0',
    '-1': '0',
    '-9999': '-1',
    '-7': '-7'
    },
  null,
  null);
  if(LTPA_dur_val != -7 && LTPA_dur_val != -1 && !LTPA_dur_val.isNull().value()){
    val_to_add = Number(LTPA_dur_val);
    LTPA_sum = LTPA_sum + val_to_add;
  } else if(LTPA_dur_val == -7) {
    LTPA_miss = LTPA_miss + 1;
  } else {
    LTPA_sum = LTPA_sum;
  }
}
if(LTPA_miss == LTPA_dur.length) {
  LTPA_sum = -1;
}
LTPA_sum;
```

Harmonisation code

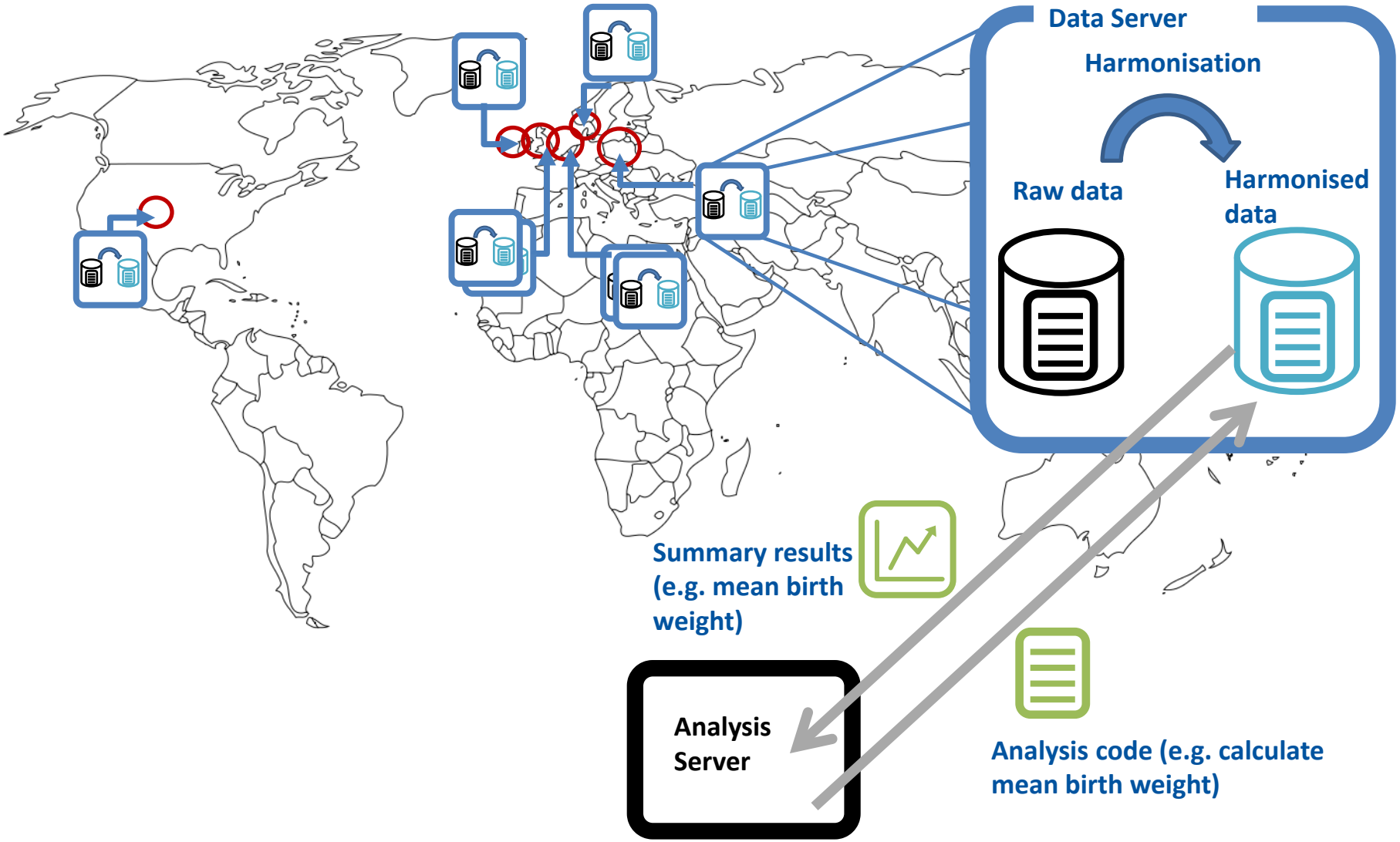


Validity of summary statistics for harmonised variables was checked

	ABCD	ALSPAC	HSS	REPRO	ROLO	SWS
Mean Birth weight (g)	3,503	3,488	3,288	3,403	4,048	3,519
Median LTPA (h/wk)	2.0	4.0	3.0	4.0	1.7	6.8

- ROLO higher birth weight as study focused on second born
- SWS higher activity due to physical activity question including additional activities compared to other studies
- Check all studies and variables before analysis starts

Seven participating studies have set up a server & harmonised data ready for analysis



InterConnect

Global data for diabetes and obesity research

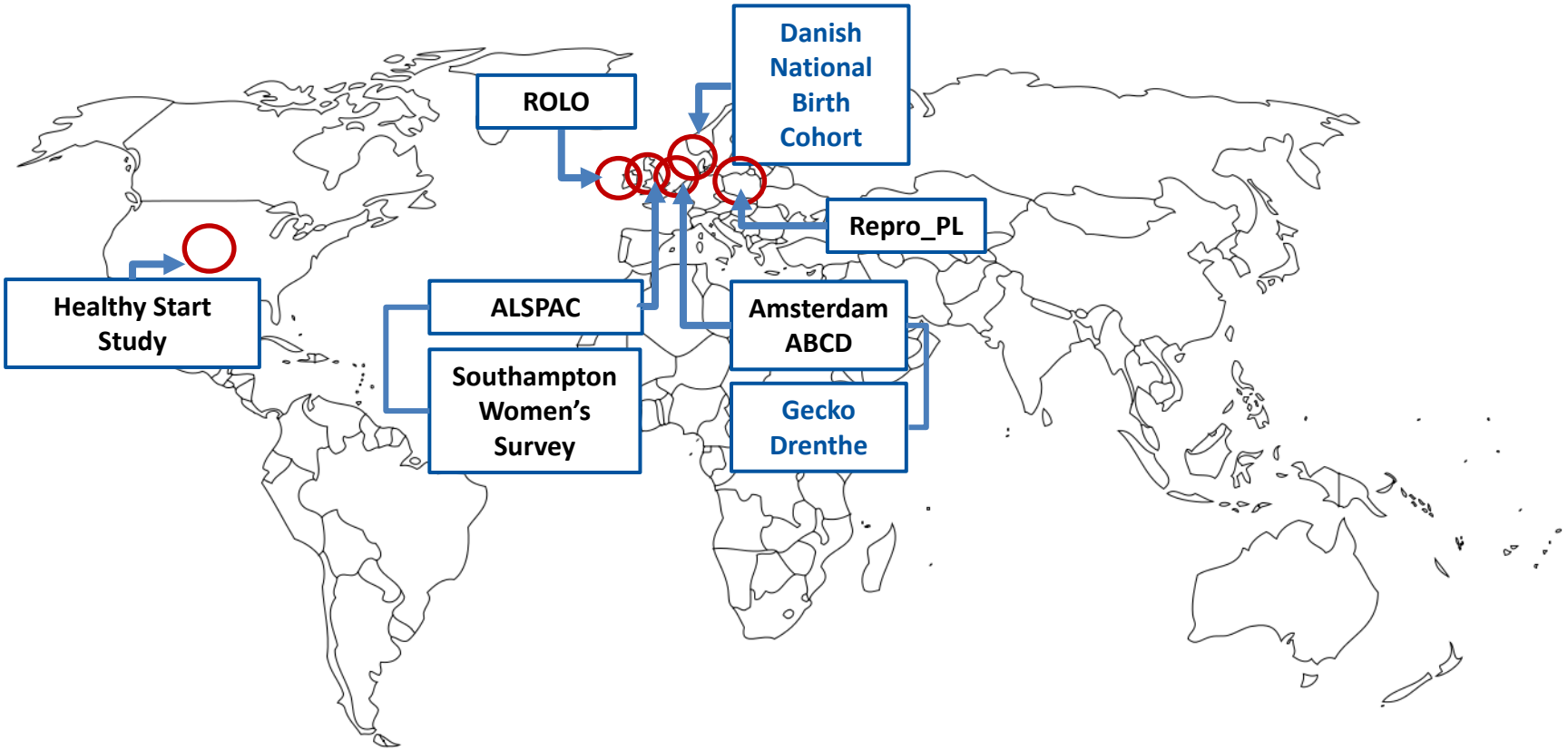
Physical activity during pregnancy exemplar project - RESULTS

Ken Ong
Programme Leader
MRC Epidemiology Unit, University of Cambridge, UK

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

Eight participating studies

(all observational)



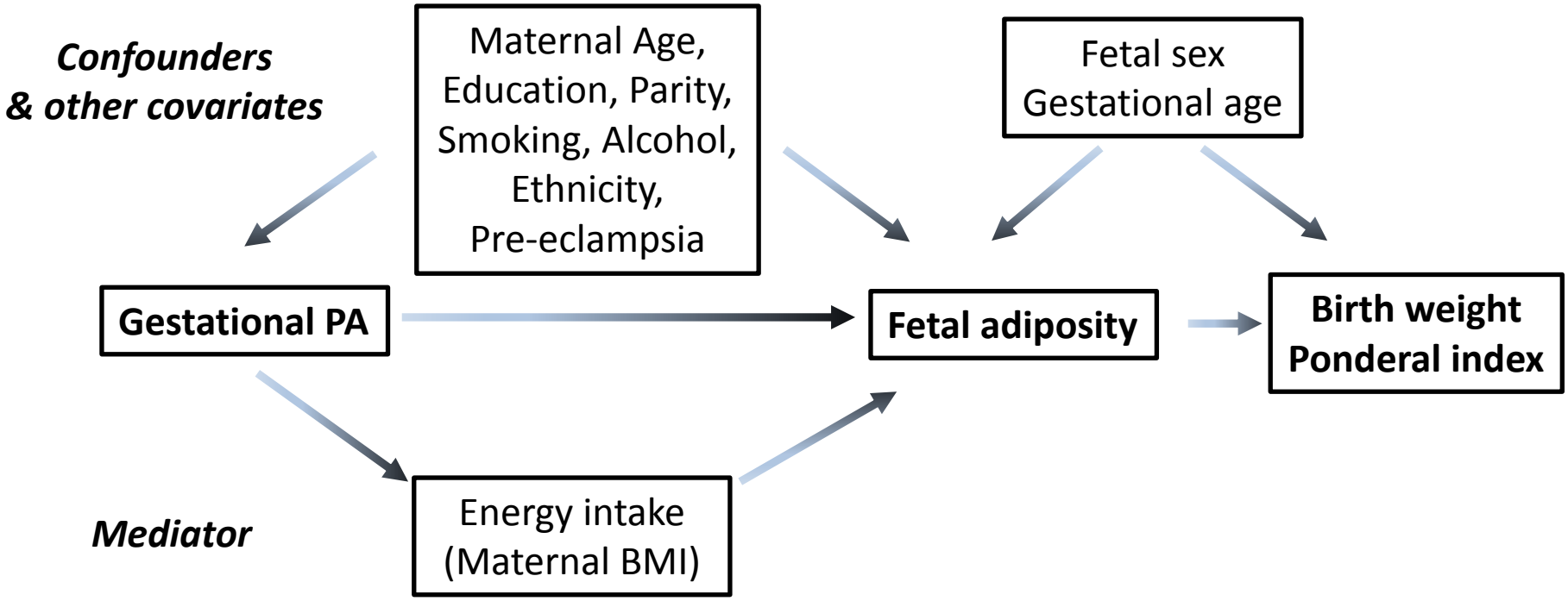
Analysis plan

- **Population:**
  - **Include:** Live births, singleton, full term babies
  - **Exclude**: Preterm (< 37 weeks gestation), multiple births

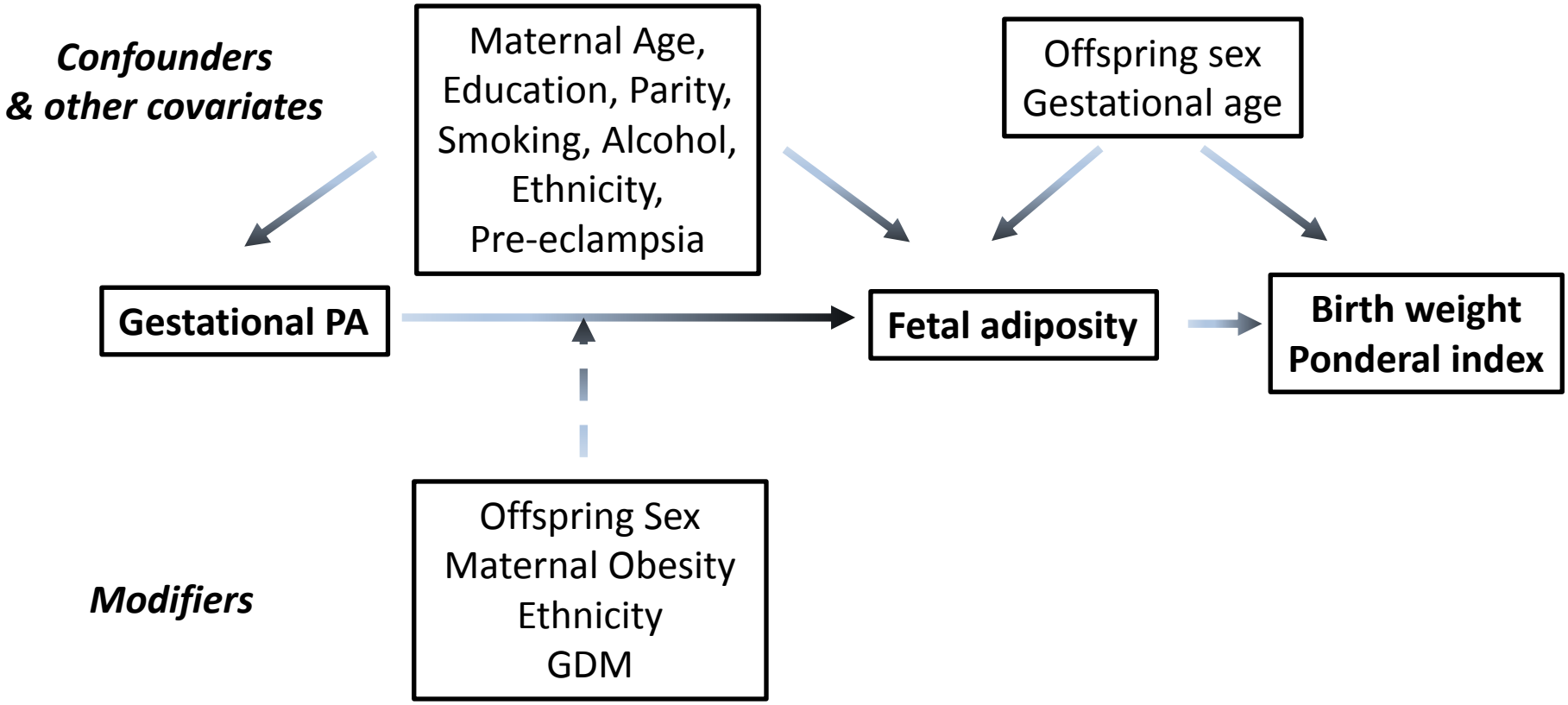
- **Exposure:** Physical activity in pregnancy: objective or subjective measurement of:
  - **Volume** (Duration or Energy Expenditure)
  - **Intensity** (Low vs. Mod/Vigorous)
  - Gestational age at PA measurement (by trimester)

- **Outcomes:**
  - **Birth weight**: Continuous or Macrosomia (Birth weight >4000 g; or LGA, large for gestational age)
  - **%body fat** in newborns (by DXA, skinfold thickness, or PeaPod air displacement plethysmography)
  - Ponderal index (BW/Length^3)

DAG for the Gestational Physical Activity – Fetal Adiposity (BW) Exemplar



DAG for the Gestational Physical Activity – Fetal Adiposity (BW) Exemplar



Study descriptions

	ABCD	ALSPAC	HSS	REPRO	ROLO	SWS
	Netherlands	UK	US	Poland	Ireland	UK
N	6464	9058	1054	991	617	1902

Study descriptions

	ABCD	ALSPAC	HSS	REPRO	ROLO	SWS
	Netherlands	UK	US	Poland	Ireland	UK
N	6464	9058	1054	991	617	1902
Non-White	**31%**	2%	**23%**	0%	2%	3%
Obese	8%	6%	**21%**	4%	**18%**	**17%**
GDM	1%	0.5%	4%	4%	2%	1%
LGA	19%	21%	37%	19%	**62%**	19%

Study descriptions

	ABCD	ALSPAC	HSS	REPRO	ROLO	SWS
	Netherlands	UK	US	Poland	Ireland	UK
N	6464	9058	1054	991	617	1902
Non-White	**31%**	2%	**23%**	0%	2%	3%
Obese	8%	6%	**21%**	4%	**18%**	**17%**
GDM	1%	0.5%	4%	4%	2%	1%
LGA	19%	21%	37%	19%	**62%**	19%
1st Trimester PA						
LTPA_Dur (h/wk)	2.0	4.0	3.0	4.0	1.7	**6.8**

Study descriptions

	ABCD	ALSPAC	HSS	REPRO	ROLO	SWS
	Netherlands	UK	US	Poland	Ireland	UK
N	6464	9058	1054	991	617	1902
Non-White	**31%**	2%	**23%**	0%	2%	3%
Obese	8%	6%	**21%**	4%	**18%**	**17%**
GDM	1%	0.5%	4%	4%	2%	1%
LGA	19%	21%	37%	19%	**62%**	19%
1st Trimester PA						
LTPA_Dur (h/wk)	2.0	4.0	3.0	4.0	1.7	**6.8**
LTPA_EE (Met.h/wk)	8.1	15.2	10.2	16.5	**4.5**	18.0
MVPA (h/wk)	1.5	4.0	1.5	4.0	**0.3**	1.3

RESULTS – Determinants of Birth weight (g)

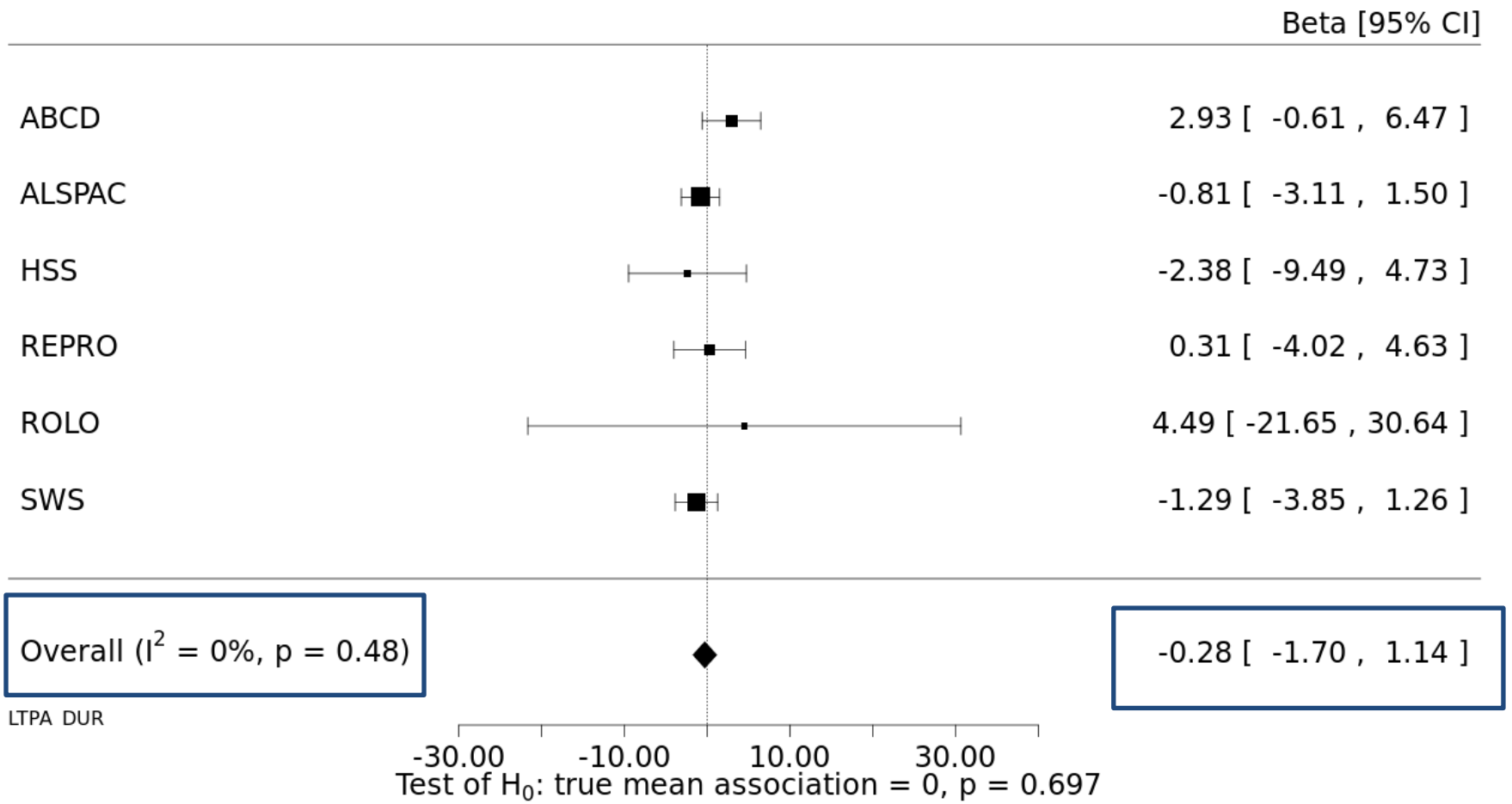
	ABCD	ALSPAC	HSS	REPRO	ROLO	SWS
	Netherlands	UK	US	Poland	Ireland	UK
N	6464	9058	1054	991	617	1902
Difference in BW						
Sex (M vs F)	138.0	139.2	164.9	173.4	167.4	158.5
Gestational age (per week)	150.5	124.2	165.4	128.0	143.7	146.1

All p<0.001

Main effects - adjusted for confounders

LEISURE TIME PA DURATION (h/wk) → BIRTHWEIGHT (g)

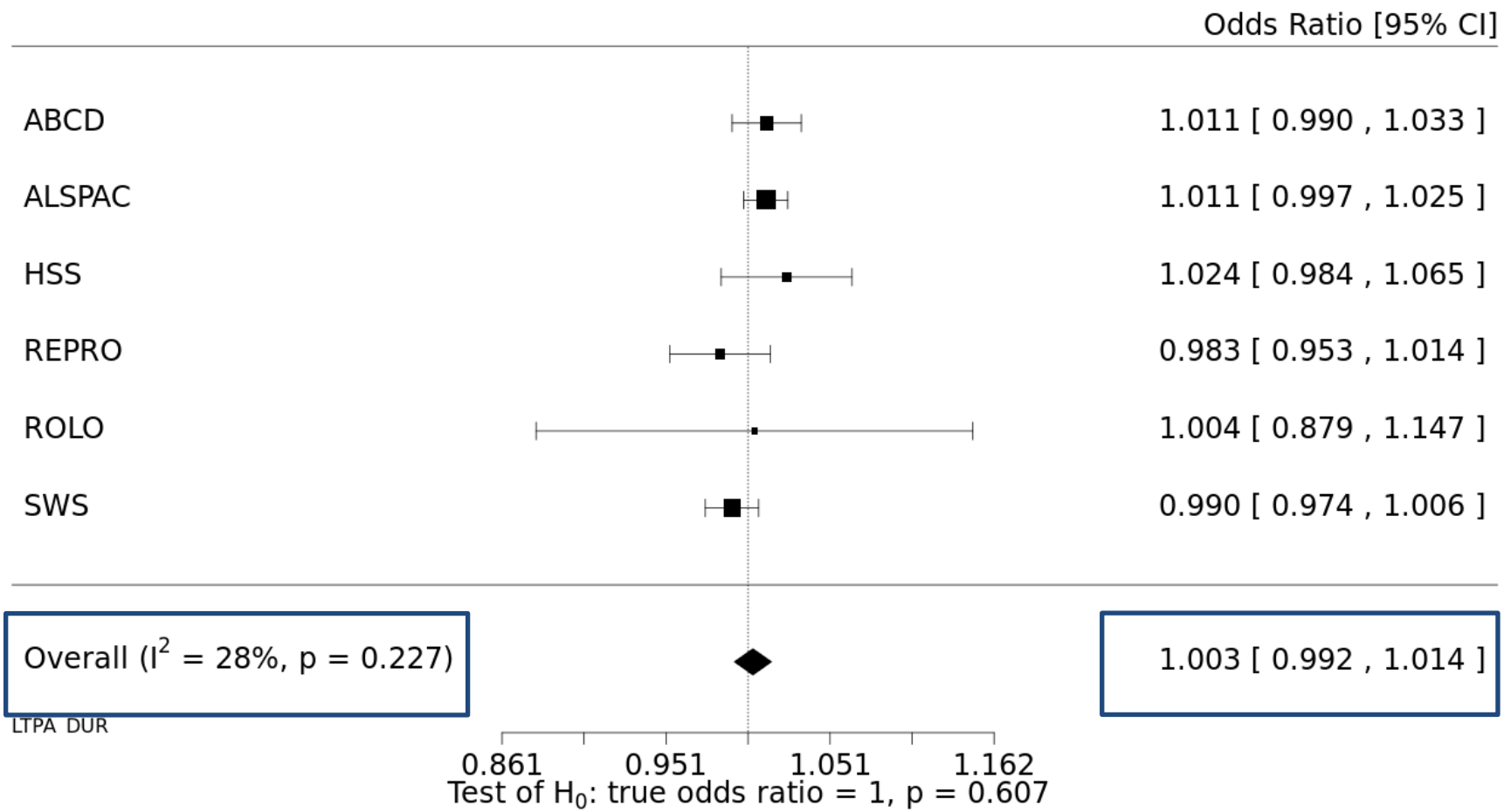
BIRTH_WEIGHT ~ LTPA_DUR + GESTATIONAL_AGE + SEX +
PARITY + MATERNAL_AGE + SMOKING + ALCOHOL + MATERNAL_EDU +
ETHNICITY



Main effects - adjusted for confounders

LEISURE TIME PA DURATION (h/wk) →LARGE FOR GESTATIONAL AGE (LGA)

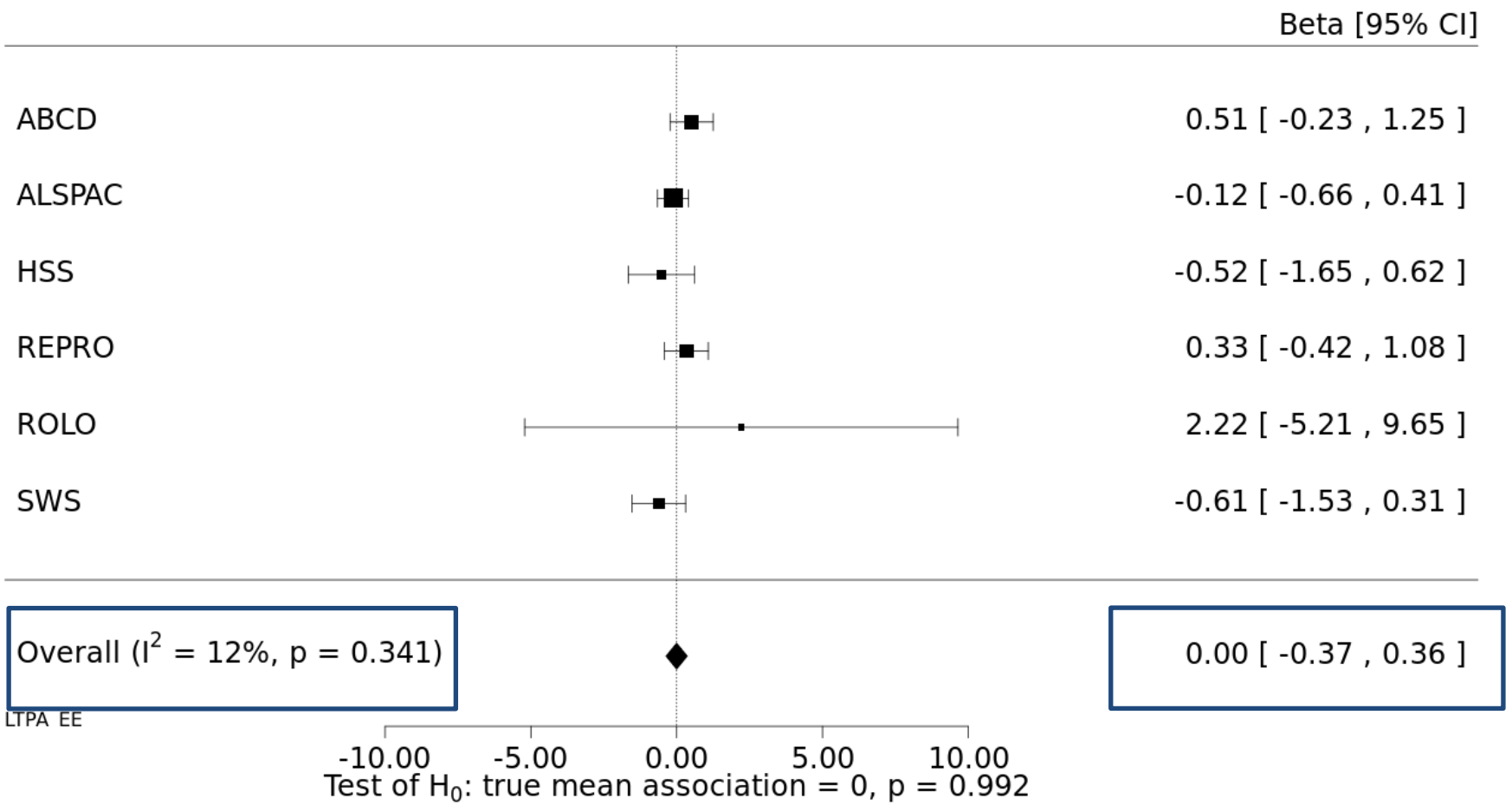
BIRTH_WEIGHT_LGA ~ LTPA_DUR + GESTATIONAL_AGE + SEX +
PARITY + MATERNAL_AGE + SMOKING + ALCOHOL + MATERNAL_EDU +
ETHNICITY



Main effects - adjusted for confounders

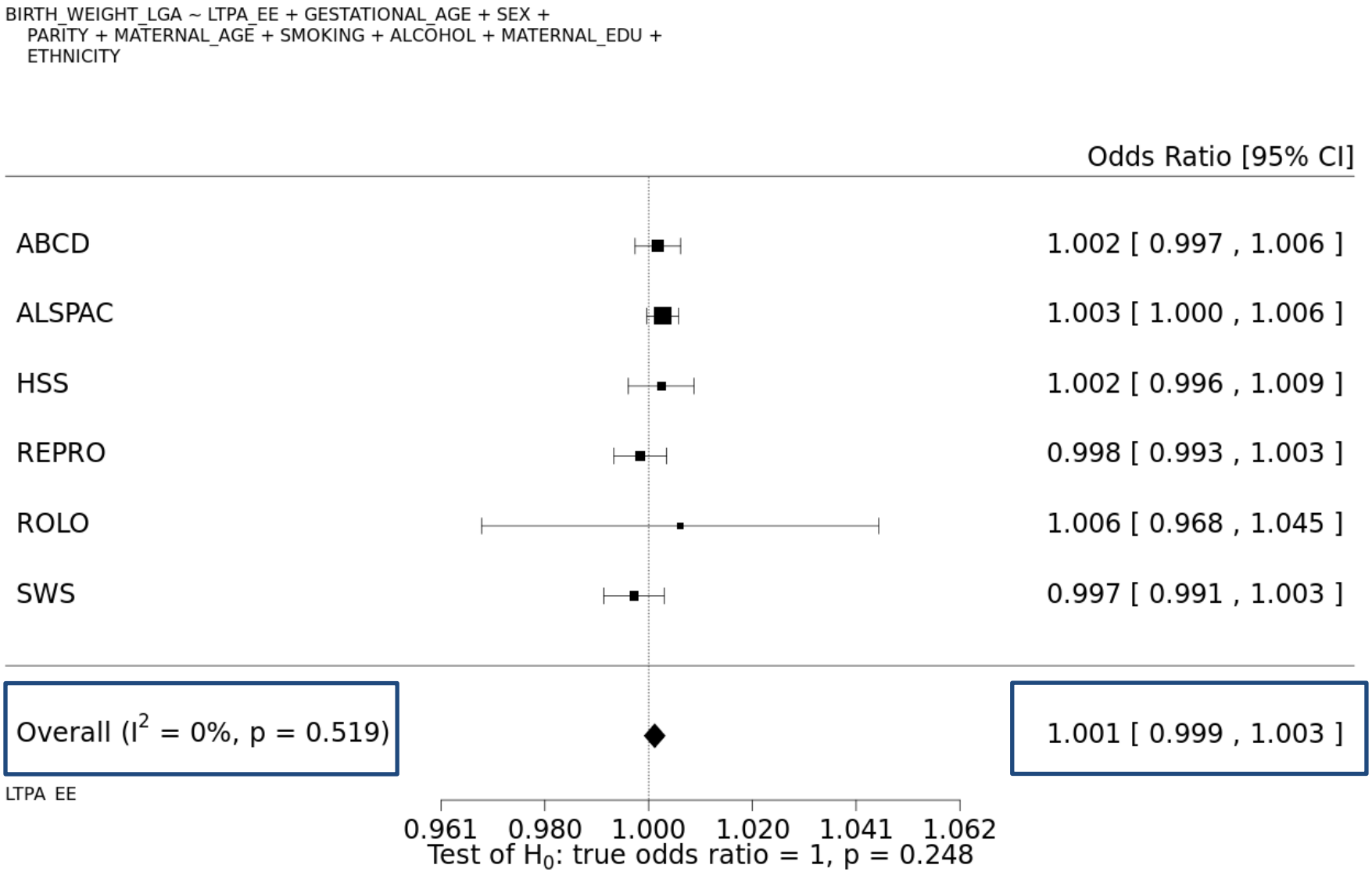
LEISURE TIME PA ENERGY EXPENDITURE (MET-h/wk) → BIRTHWEIGHT (g)

BIRTH_WEIGHT ~ LTPA_EE + GESTATIONAL_AGE + SEX +
PARITY + MATERNAL_AGE + SMOKING + ALCOHOL + MATERNAL_EDU +
ETHNICITY



Main effects - adjusted for confounders

LEISURE TIME PA ENERGY EXPENDITURE (MET-h/wk) → LGA



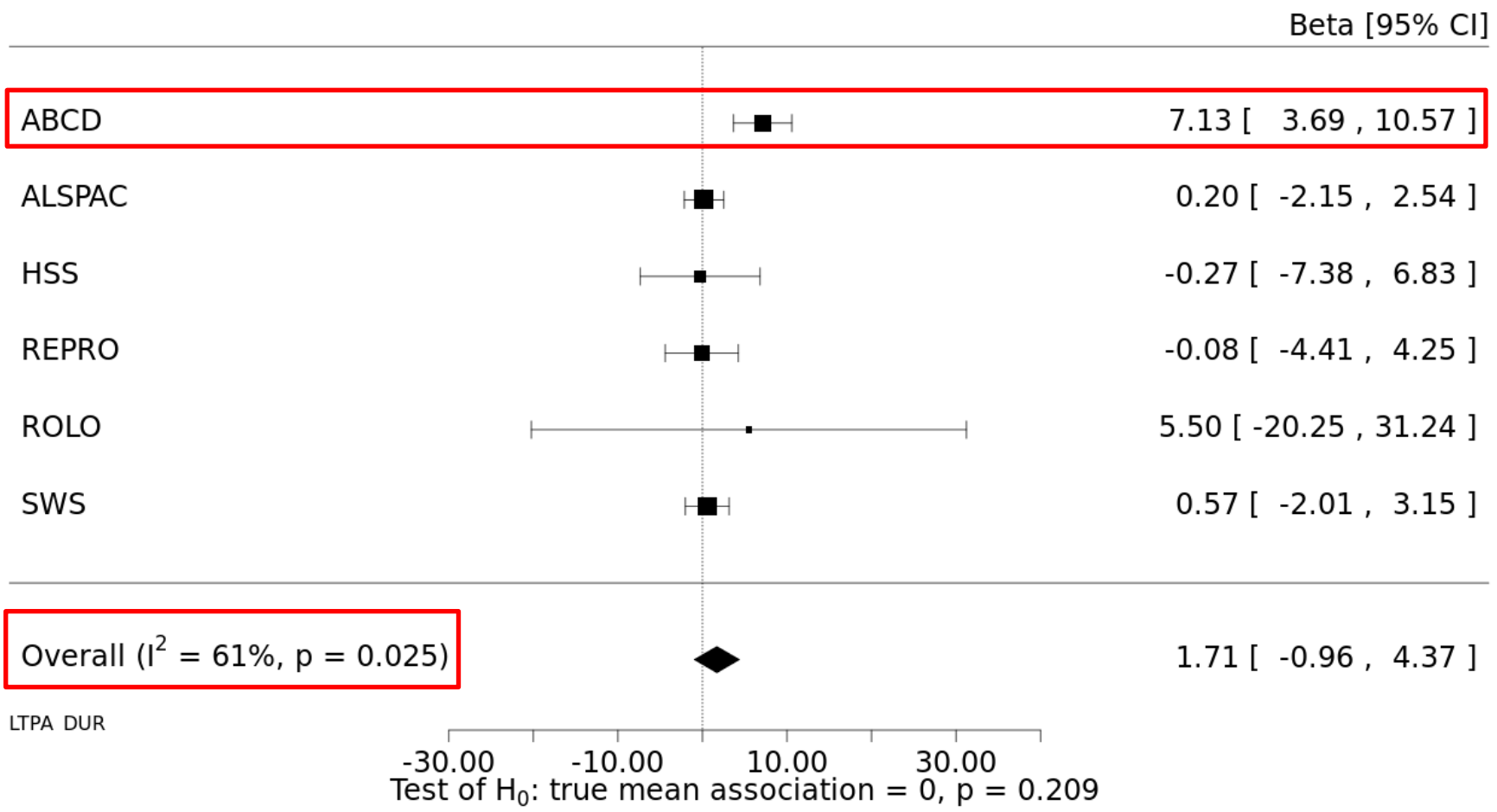
Explore the effect of adjustment for confounding

- Tested models **WITH** and **WITHOUT** adjustment for potential confounders (Maternal Education, Age, Parity, Smoking, Alcohol, Ethnicity)

Main effects – WITHOUT adjustment for confounders

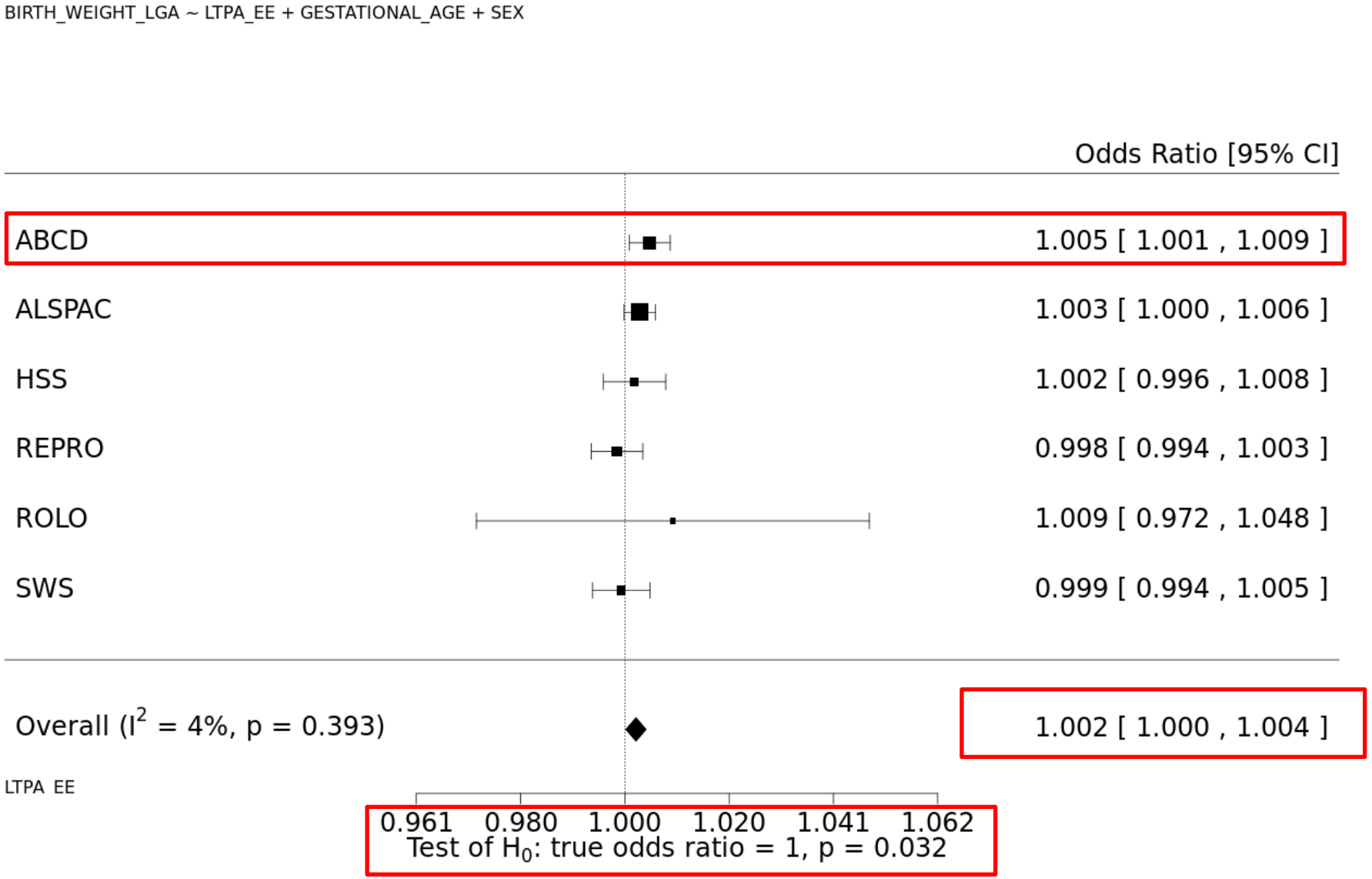
LEISURE TIME PA DURATION (h/wk) → BIRTHWEIGHT (g)

BIRTH_WEIGHT ~ LTPA_DUR + GESTATIONAL_AGE + SEX



Main effects – WITHOUT adjustment for confounders

LEISURE TIME PA ENERGY EXPENDITURE (MET-h/wk) → LGA



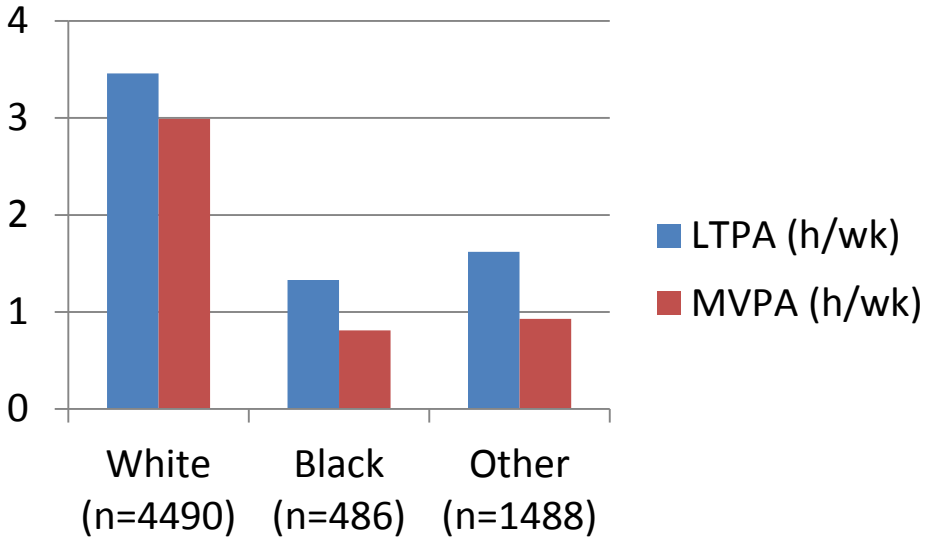
ABCD Study

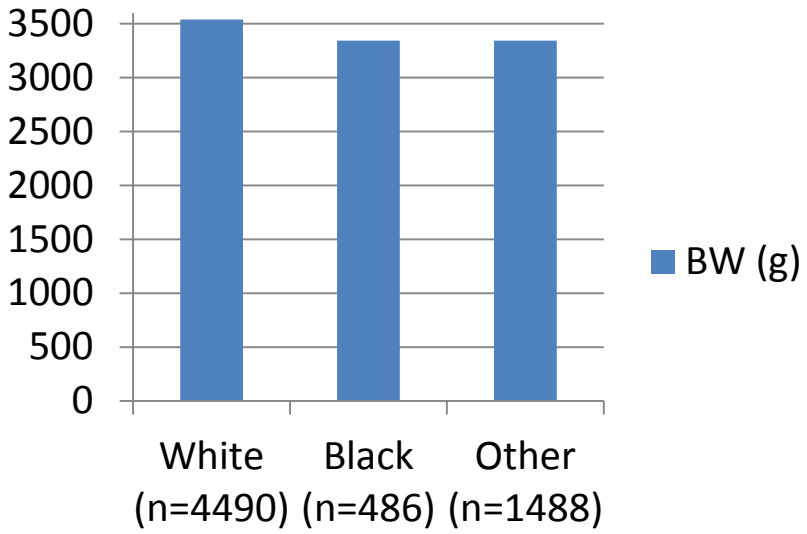
LTPA Duration (h/wk) → Birthweight (g)

Model Covariates	Estimate	Std. Error	P-value
GESTATIONAL_AGE+SEX	7.16	1.75	4.4 E-05
GESTATIONAL_AGE+SEX+PARITY	9.02	1.74	2.0 E-07
GESTATIONAL_AGE+SEX+PARITY+MATERNAL_AGE	7.54	1.77	2.1 E-05
GESTATIONAL_AGE+SEX+PARITY+MATERNAL_AGE +SMOKING	6.97	1.76	7.7 E-05
GESTATIONAL_AGE+SEX+PARITY+MATERNAL_AGE +SMOKING+ALCOHOL	6.97	1.76	7.7 E-05
GESTATIONAL_AGE+SEX+PARITY+MATERNAL_AGE +SMOKING+ALCOHOL+**MATERNAL_EDUCATION**	5.38	1.79	2.6 E-03
GESTATIONAL_AGE+SEX+PARITY+MATERNAL_AGE +SMOKING+ALCOHOL+**MATERNAL_EDUCATION** +**ETHNICITY**	2.94	1.80	1.0 E-01

ABCD Study

Physical activity and BW vary markedly by Ethnicity





Explore the effect of adjustment for confounding

- Tested models **BEFORE** and **AFTER** adjustment for potential confounders (Maternal SES, Age, Parity, Smoking, Alcohol, Ethnicity)
- Adjustment for confounding **reduced heterogeneity** and **reduced positive confounding** due to Education and Ethnicity

Interactions with **Offspring Sex (male/female)**

Association tested	Interaction Estimate	95% CI	*P-value*	I-square
LTPA Duration (h/wk) → BW (g)	-0.5	-3.1 to 2.1	*0.7*	0%
LTPA EE (MetH/wk) → BW (g)	-0.2	-0.8 to 0.4	*0.6*	0%
MVPA (h/wk) → BW (g)	-0.6	-3.9 to 2.7	*0.7*	0%
LTPA Duration (h/wk) → LGA	1.00	0.97 to 1.02	*0.8*	34%
LTPA EE (MetH/wk) → LGA	1.00	0.99 to 1.00	*0.2*	0%
MVPA (h/wk) → LGA	0.99	0.97 to 1.02	*0.6*	28%

Interactions with **Ethnicity (White/Black)**

Association tested	Interaction Estimate	95% CI	*P-value*	I-square
LTPA Duration (h/wk) → BW (g)	0.5	-10.2 to 11.2	*0.9*	0%
LTPA EE (MetH/wk) → BW (g)	0.5	-1.9 to 2.9	*0.7*	0%
MVPA (h/wk) → BW (g)	-2.7	-15.6 to 10.2	*0.7*	0%
LTPA Duration (h/wk) → LGA	0.98	0.89 to 1.07	*0.6*	26%
LTPA EE (MetH/wk) → LGA	1.00	0.98 to 1.02	*0.9*	26%
MVPA (h/wk) → LGA	0.99	0.87 to 1.11	*0.8*	38%

Interactions with **Maternal Obesity (no/yes)**

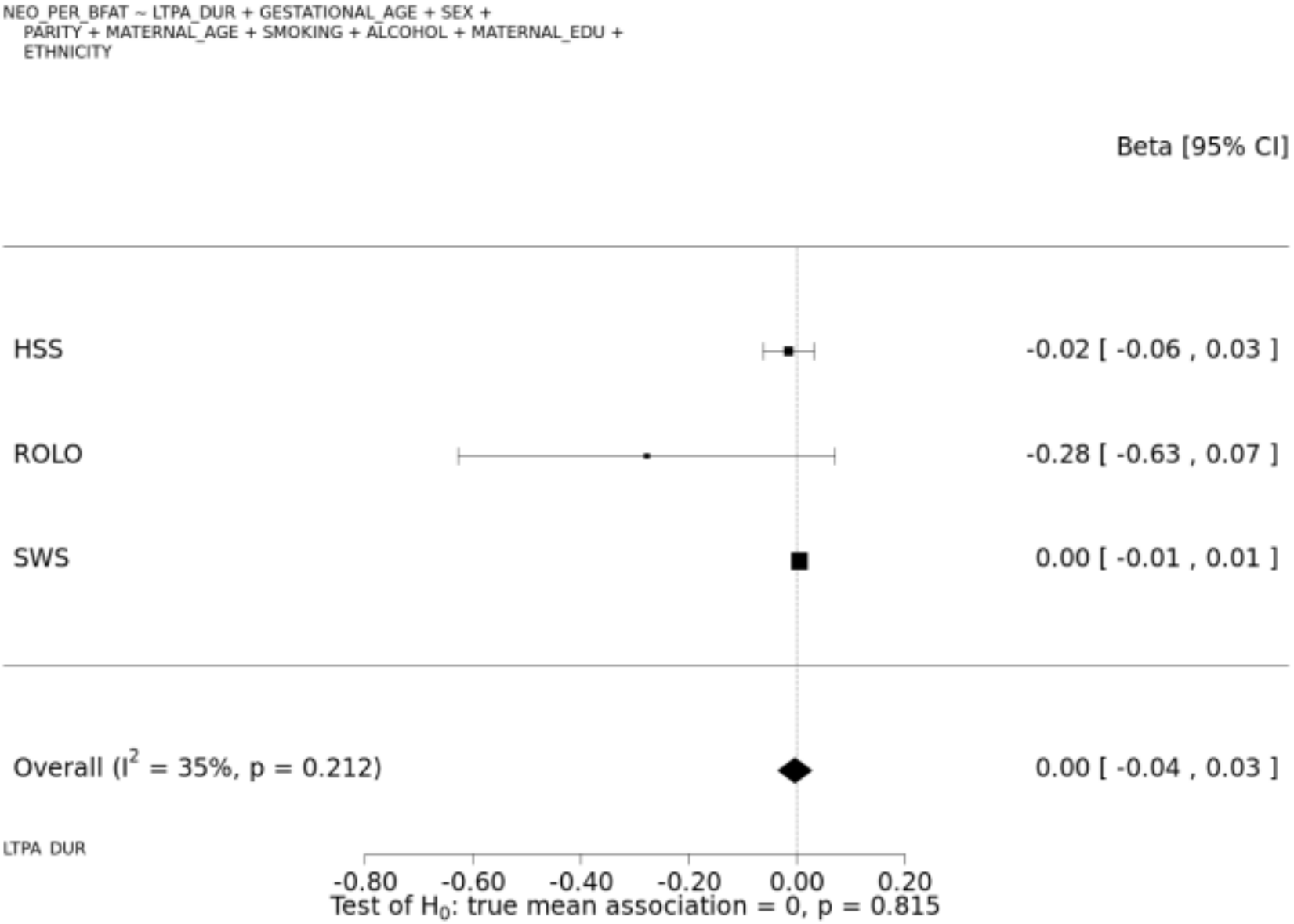
Association tested	Interaction Estimate	95% CI	*P-value*	I-square
LTPA Duration (h/wk) → BW (g)	1.3	-2.6 to 5.3	*0.5*	0%
LTPA EE (MetH/wk) → BW (g)	0.2	-0.9 to 1.2	*0.8*	0%
MVPA (h/wk) → BW (g)	-0.8	-6.5 to 5.0	*0.8*	0%
LTPA Duration (h/wk) → LGA	0.99	0.97 to 1.02	*0.5*	0%
LTPA EE (MetH/wk) → LGA	1.00	0.99 to 1.00	*0.3*	0%
MVPA (h/wk) → LGA	0.99	0.96 to 1.03	*0.7*	0%

Interactions with **GDM (no/yes)**

Association tested	Interaction Estimate	95% CI	*P-value*	I-square
LTPA Duration (h/wk) → BW (g)	5.3	-13.8 to 24.3	*0.6*	46%
LTPA EE (MetH/wk) → BW (g)	2.4	-2.3 to 7.0	*0.3*	55%
MVPA (h/wk) → BW (g)	10.0	-17.9 to 37.9	*0.5*	56%
LTPA Duration (h/wk) → LGA	1.00	0.93 to 1.08	*1.0*	0%
LTPA EE (MetH/wk) → LGA	1.01	0.99 to 1.03	*0.6*	0%
MVPA (h/wk) → LGA	1.01	0.90 to 1.14	*0.8*	17%

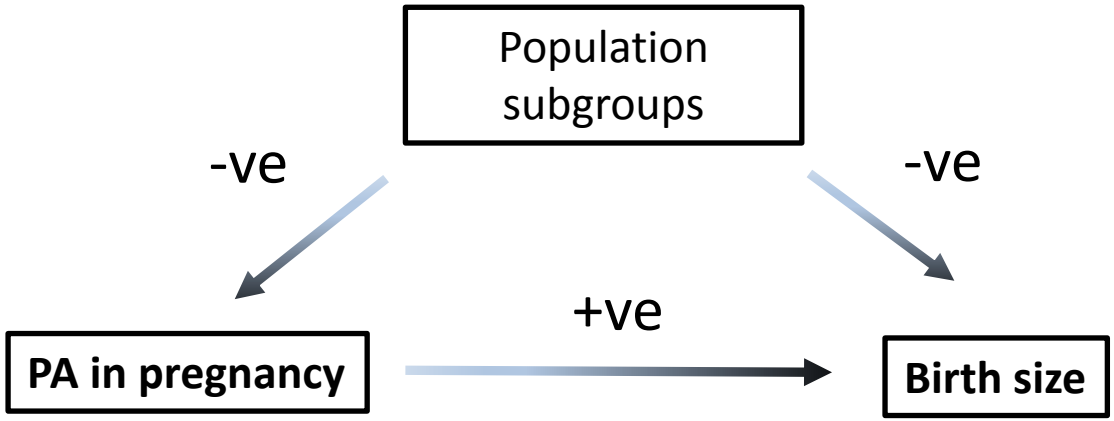
Main effects - adjusted for confounders

LEISURE TIME PA Duration (h/wk) → NEWBORN BODY FAT %



Summary

- Heterogeneity between studies & Positive Confounding – were reduced by consistent adjustment for confounders across studies



- Consistently Null association with 1st trimester PA
- No heterogeneity detectable by strata

Next Steps

- Update the current analyses
  - Ponderal index
  - Add data from DNBC (n=79K)
- Test the shape of association / possible threshold effects
  - Is a challenge to harmonise absolute values of PA
- Test associations with 3rd trimester Physical Activity
  - Five of eight studies

Programme

14:30	**The InterConnect Project**	
14:45	**Applying the InterConnect approach for federated meta-analysis** 1. Physical activity in pregnancy & neonatal anthropometric outcomes 2. Fish intake and risk of type 2 diabetes - *Why this question is important* - *Why federated meta-analysis is required* - *Progress with set up and harmonisation*	Nita Forouhi & Silvia Pastorino
16.00	**Future perspectives**	
16:30	**Discussion and involvement**	

InterConnect

Global data for diabetes and obesity research

Exemplar: Fish intake and new-onset type 2 diabetes

Silvia Pastorino, Nita Forouhi
MRC Epidemiology Unit
12th September 2016, Munich

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

What is known

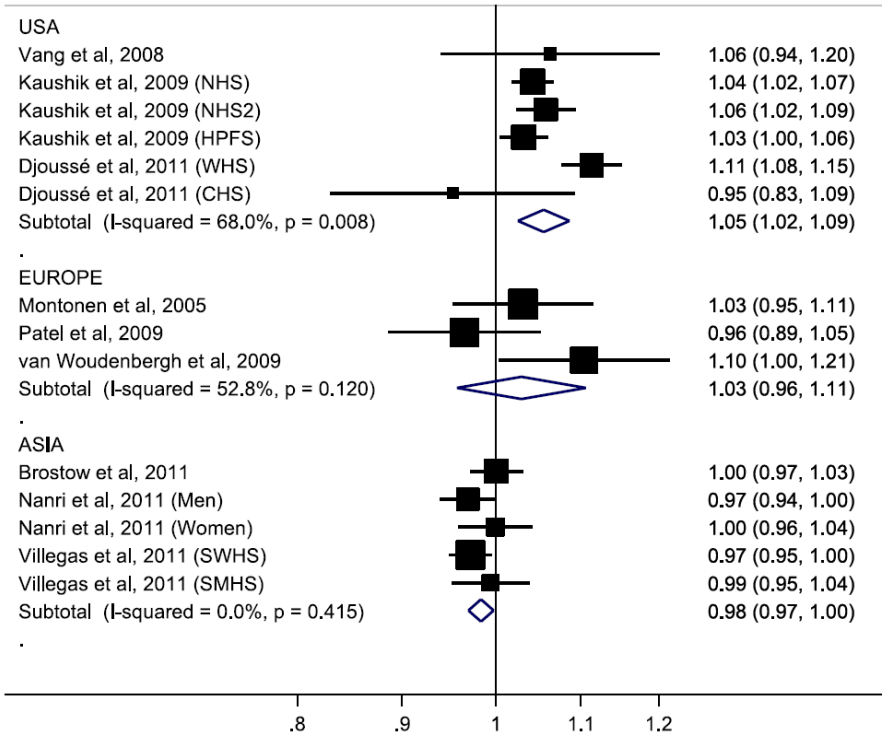
Research question

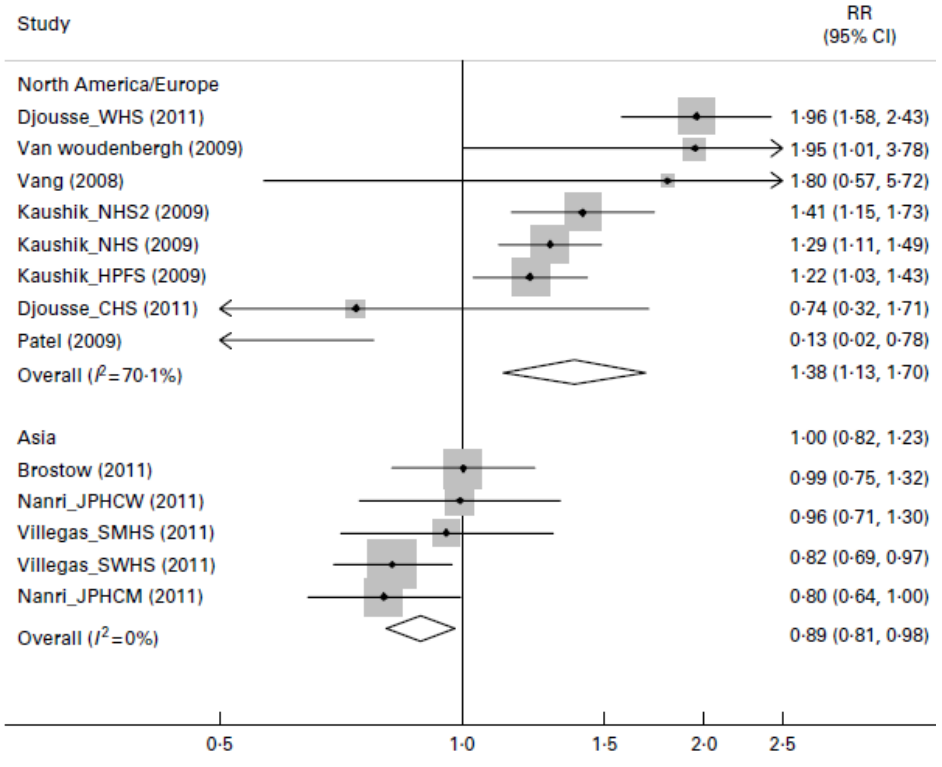
- It is proposed that fish intake is likely to be beneficial for the prevention of type 2 diabetes, based on the benefits for cardiovascular health

Meta analyses of fish and type 2 diabetes

Meta analyses	Wallin, Diabetes Care, 2012,	Wu, Br J Nutr, 2012	Xun, Diabetes Care, 2012	Zhou, Br J Nutr, 2012	Zheng, PLOS ONE, 2012
	13 studies 21,173 T2D	13 studies 20,830 T2D	12 studies 18,711 T2D	9 studies 18,272 T2D	11 studies 18,047 T2D
Overall Relative risk (95% CI)	1.01 0.99, 1.03 Per serving/week	1.12 0.94, 1.34 Per 100g/day	1.00 0.85, 1.18 Highest/lowest	1.15 0.98, 1.35 Highest/lowest	1.07 0.91, 1.25 Highest/lowest

Fish and T2D: Location matters





Per serving/week

Wallin A *Diabetes Care*, 2012, 35:

Per 100 g/d

Wu HY *BJN* 2012, 107:

What are the research gaps?

Research question

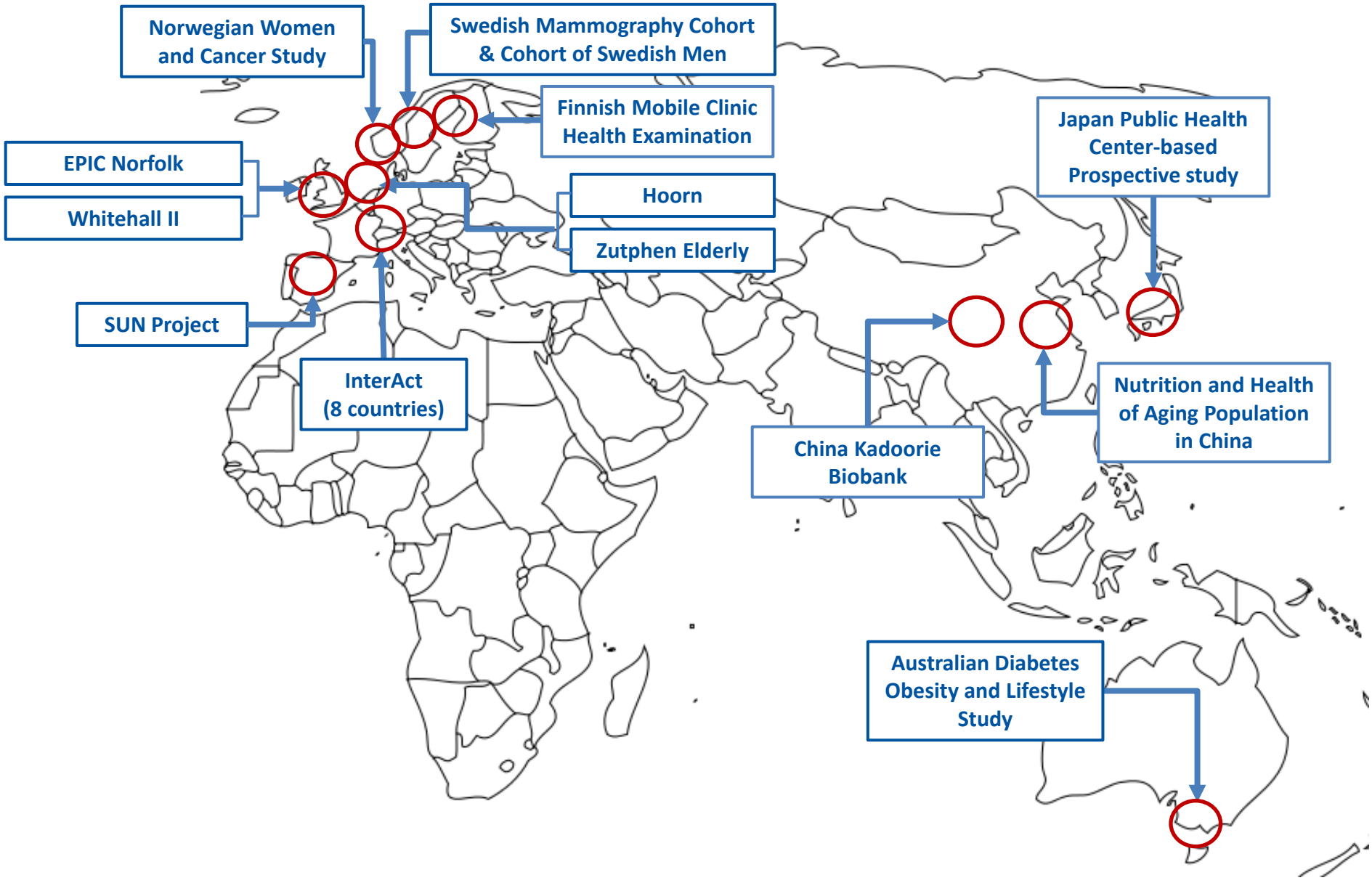
- Systematic reviews analysed total fish and did not distinguish between types of fish (e.g. fatty fish, lean fish and shellfish) or cooking methods; contaminants/pollutants might also contribute
- Systematic reviews did not include unpublished results
- High heterogeneity in meta-analyses might be caused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

Advantages of InterConnect

Research question

- Individual participant meta-analysis without physical pooling of data
- Include studies that have not yet published on the association between fish and T2D
- Can perform sub-group analyses and analyses of different fish types
- Reduce heterogeneity by:
  - Including the same types of confounders
  - Harmonising exposures and outcome to a common format

Map of participating studies



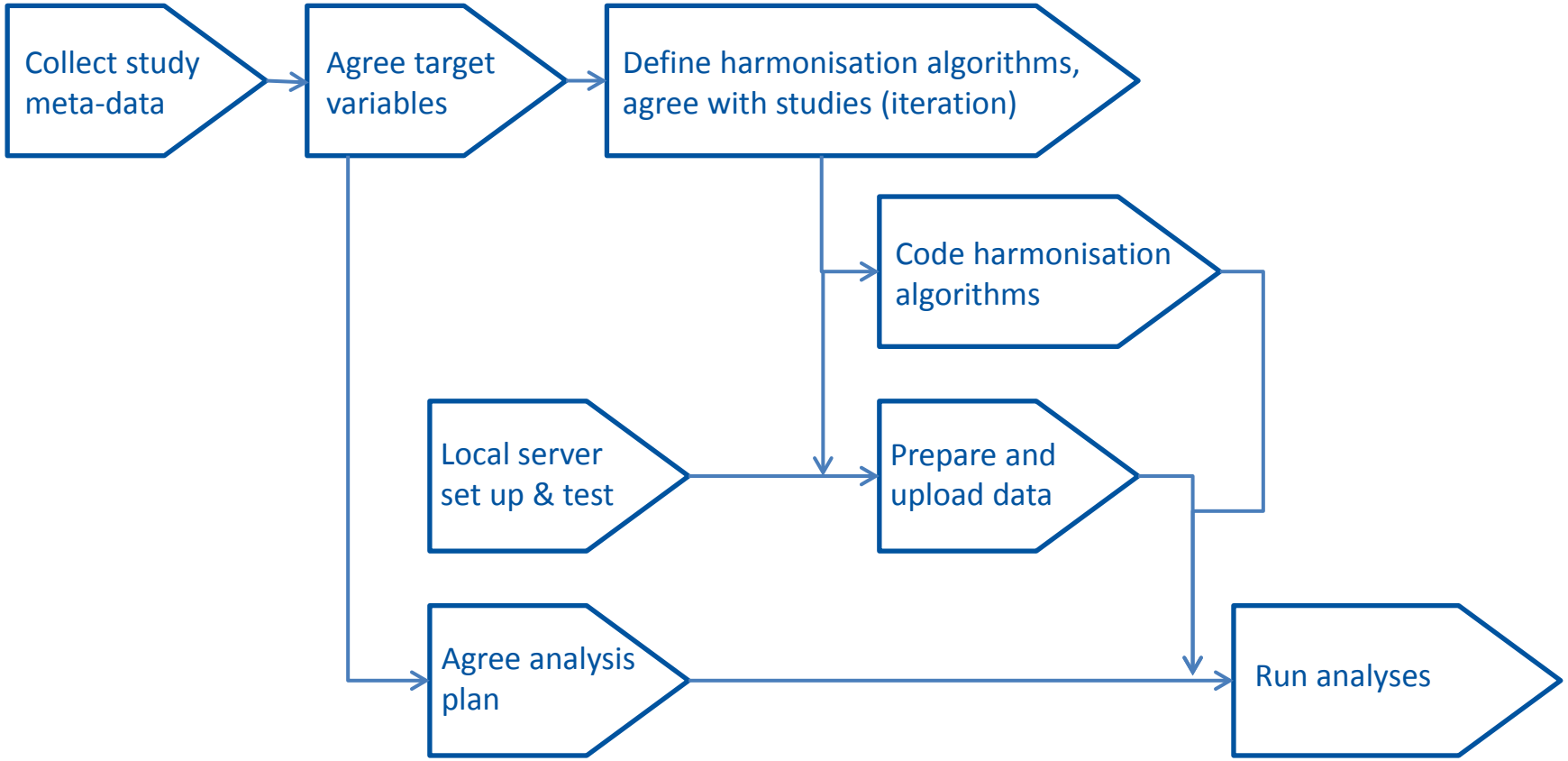
Participating studies: Europe

Study name	Country	N, sex	Self-reported method	N of fish variables
EPIC-InterAct	8 European countries	28,460 m/w	FFQ	12
EPIC Norfolk	UK	25,639 m/w	FFQ, 7-d diet diary	6
Finnish Mobile Clinic Health Examination (FMC)	Finland	4,304 m/w	Dietary history	29
Hoorn Study	Netherlands	6000 m/w	FFQ	2
Norwegian Women and Cancer Study (NOWAC)	Norway	33,740 w	FFQ	22
Swedish Mammography Cohort (SMC) & Cohort of Swedish Men (SMC, COSM)	Sweden	66,651 w & 45,906 m	FFQ	13
SUN Project	Spain	22,340 m/w	FFQ	7
Whitehall II	UK	10,308 m/w	FFQ	5
Zutphen Elderly	Netherlands	876 m/w	Cross-check dietary history	~30

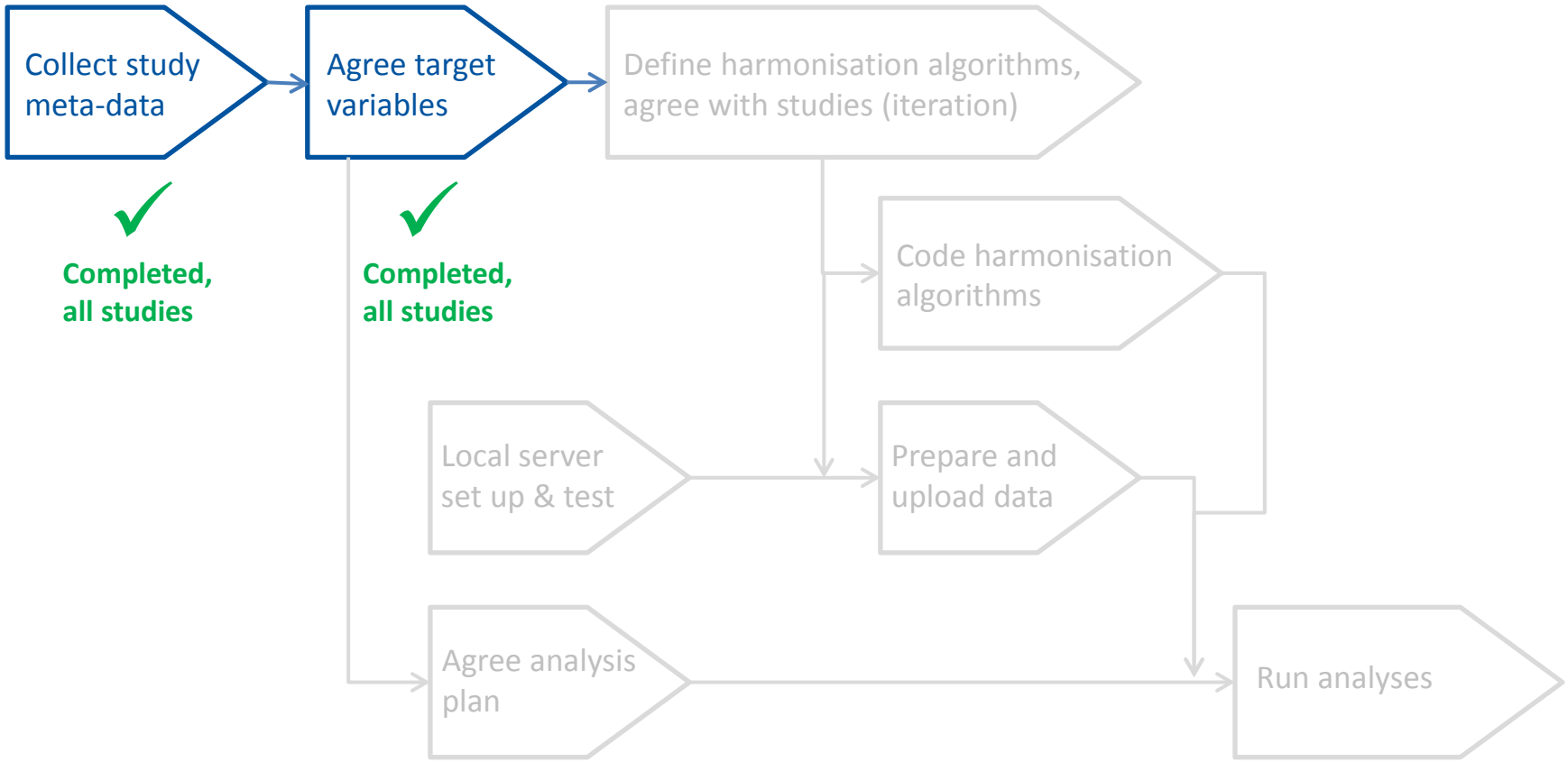
Participating studies: Asia and Australia

Study name	Country	N, sex	Self-reported method	N of fish variables
The Australian Diabetes Obesity and Lifestyle Study (AusDiab)	Australia	6537 m/w	FFQ	3
Japan Public Health Center-based Prospective study (JPHC)	Japan	52,680 m/w	FFQ	19
Nutrition and Health of Aging Population in China	China	4,526 m/w	open-ended FFQ	7
China Kadoorie Biobank	China	>500,000 m/w	FFQ	1

Overview of work-flow



Progress


Collect study meta-data
Agree target variables
Define harmonisation algorithms, agree with studies (iteration)
Completed, all studies
Completed, all studies
Code harmonisation algorithms
Local server set up & test
Prepare and upload data
Agree analysis plan
Run analyses

Target exposure harmonisation variables

- Total fish
- Fatty/oily fish (EPIC classification: fat content > 4%)
- Lean fish
- Shellfish (crustaceans and molluscs)
- Saltwater fish
- Freshwater fish
- Fried fish
- Smoked or salted fish

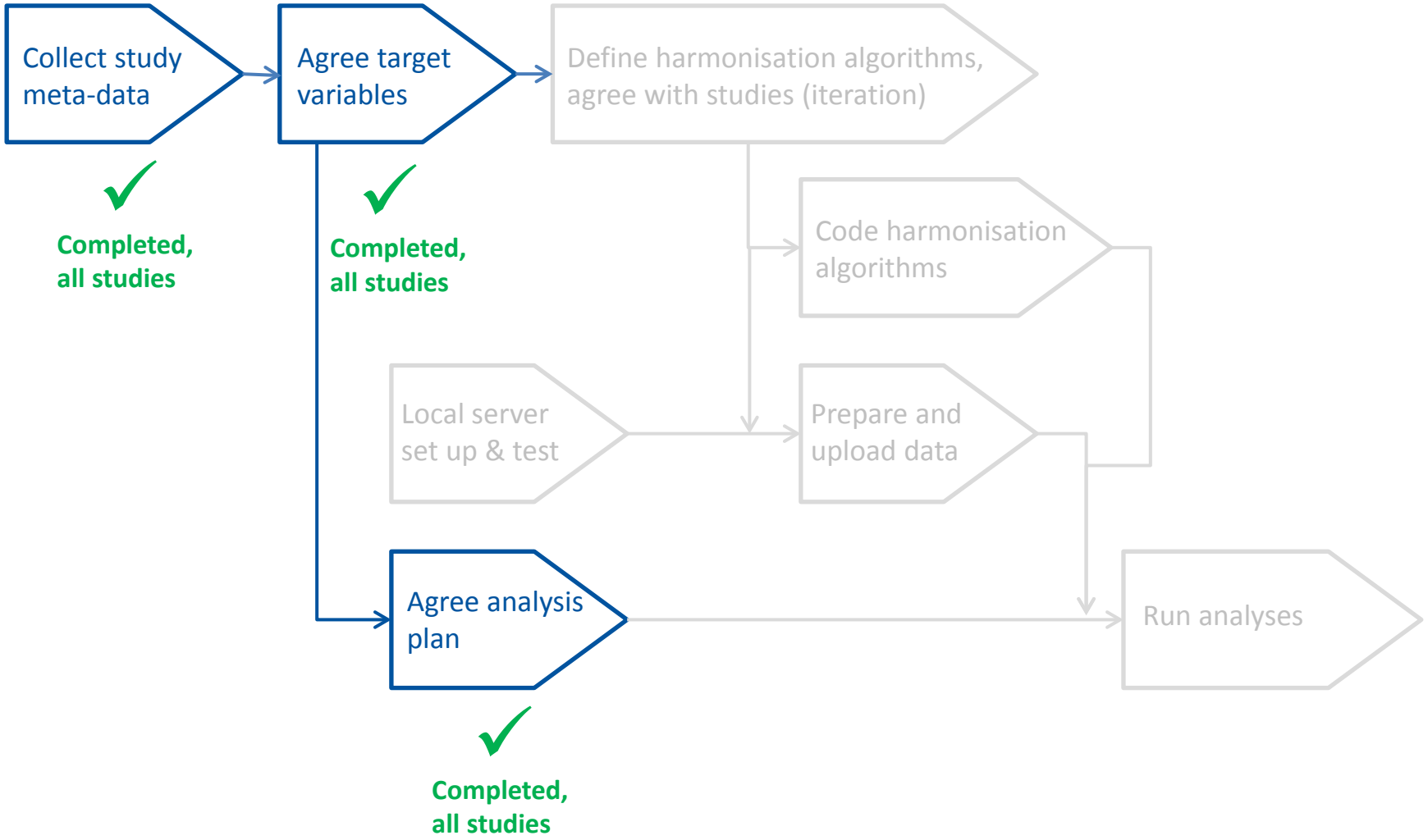
Units:

  - g/day
  - Servings /day

Fish exposure harmonisation potential

Participating cohorts	Target variables						
	Total fish	Fatty fish	Lean fish	Shellfish	Freshwater and Saltwater fish	Fried fish	Smoked/ salted fish
AusDiab	Y					Y	
EPIC Norfolk	Y	Y	Y	Y	Y	Y	
FMC	Y	Y	Y	Y	Y		Y
Hoorn	Y	Y					
EPIC-InterAct	Y	Y	Y	Y		Y	
JPHC	Y	Y	Y	Y	Y		Y
NOWAC	Y	Y	Y	Y	Y		
Nutrition and Health of Aging Population in China	Y			Y	Y		Y
SMC and COSM	Y	Y	Y	Y		Y	Y
SUN	Y	Y	Y	Y			Y
Whitehall II	Y	Y	Y	Y		Y	
Zutphen Elderly	Y	Y	Y	Y	Y	Y	?
China Kadoorie Biobank	y						

Progress

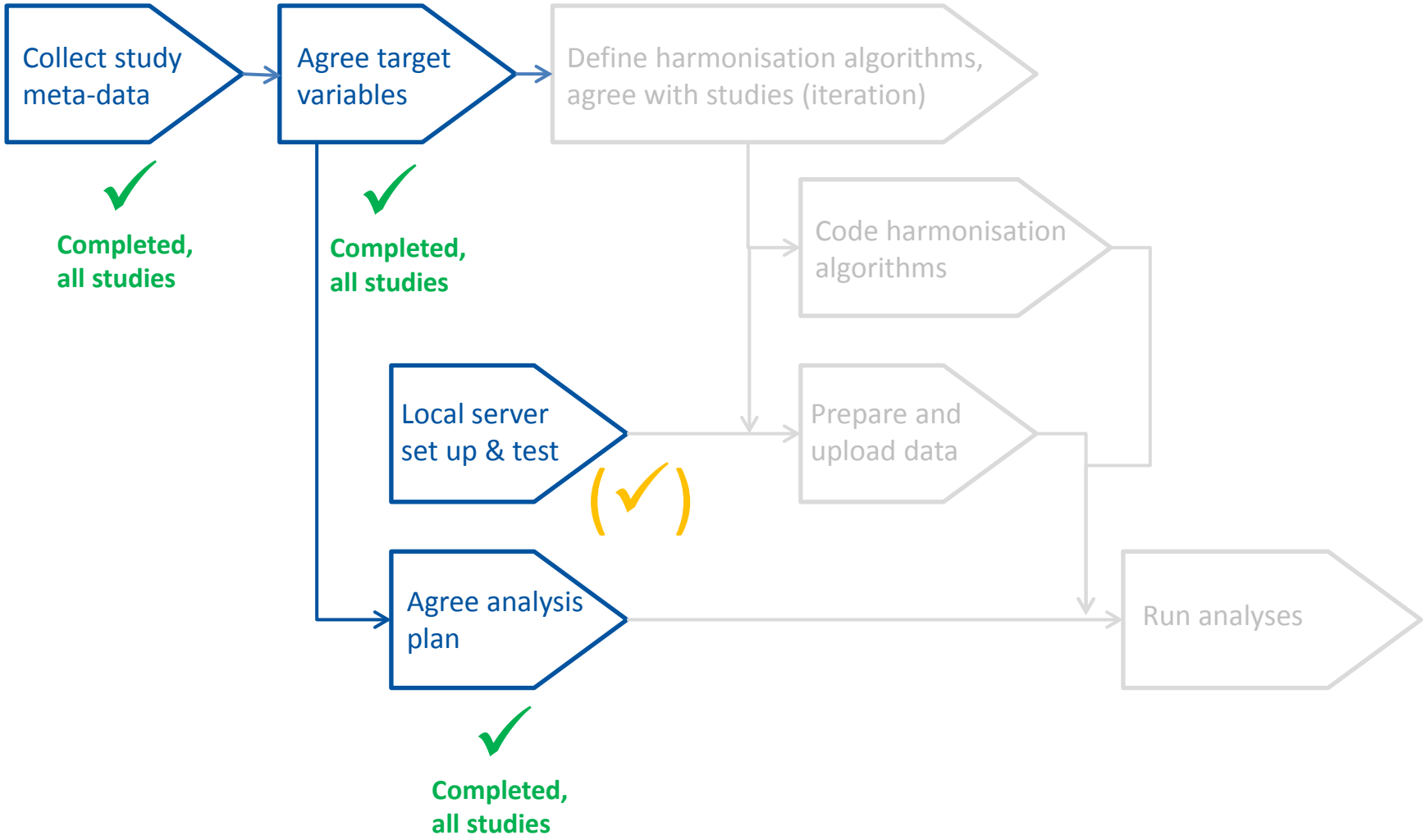

Collect study meta-data
Agree target variables
Define harmonisation algorithms, agree with studies (iteration)
Completed, all studies
Completed, all studies
Code harmonisation algorithms
Local server set up & test
Prepare and upload data
Agree analysis plan
Run analyses
Completed, all studies

Analysis plan

- **Population:** Include general population, no limits on age, exclude prevalent cases of type 2 diabetes
- **Exposure:** Fish consumption: at least one measure of frequency (servings/day or servings/week) or quantity (g/day or per serving) of intake of total and/or type of fish/seafood consumed
- **Outcome:** Type 2 diabetes incidence: biochemically diagnosed or self-reported (ideally has information on diagnosis date, and ideally validated against an additional source)
- **Potential Confounders:**
  - **Demographic:** age, sex, socio-economic status (education and/or occupation, income if available);
  - **Lifestyle:** smoking, physical activity, alcohol;
  - **Health:** BMI (waist circumference or waist-hip ratio if available); family history of diabetes, co-morbidity (MI, stroke, cancer, hypertension)*;
  - **Dietary:** total energy intake, other dietary variables (other key food groups – eg red/processed meat, fruits/vegetables, dairy, sugary beverages, fibre), cooking method (information on frying, grilling, eaten raw, etc), cooking oil or other fat used in cooking, supplements (particularly fish-oil supplement).
- **Modifiers:** Geographic location (continent, region), environmental contaminants (if available) such as PCBs, methyl mercury, dioxins and others relevant compounds.

* *Or exclude people with prevalent conditions*

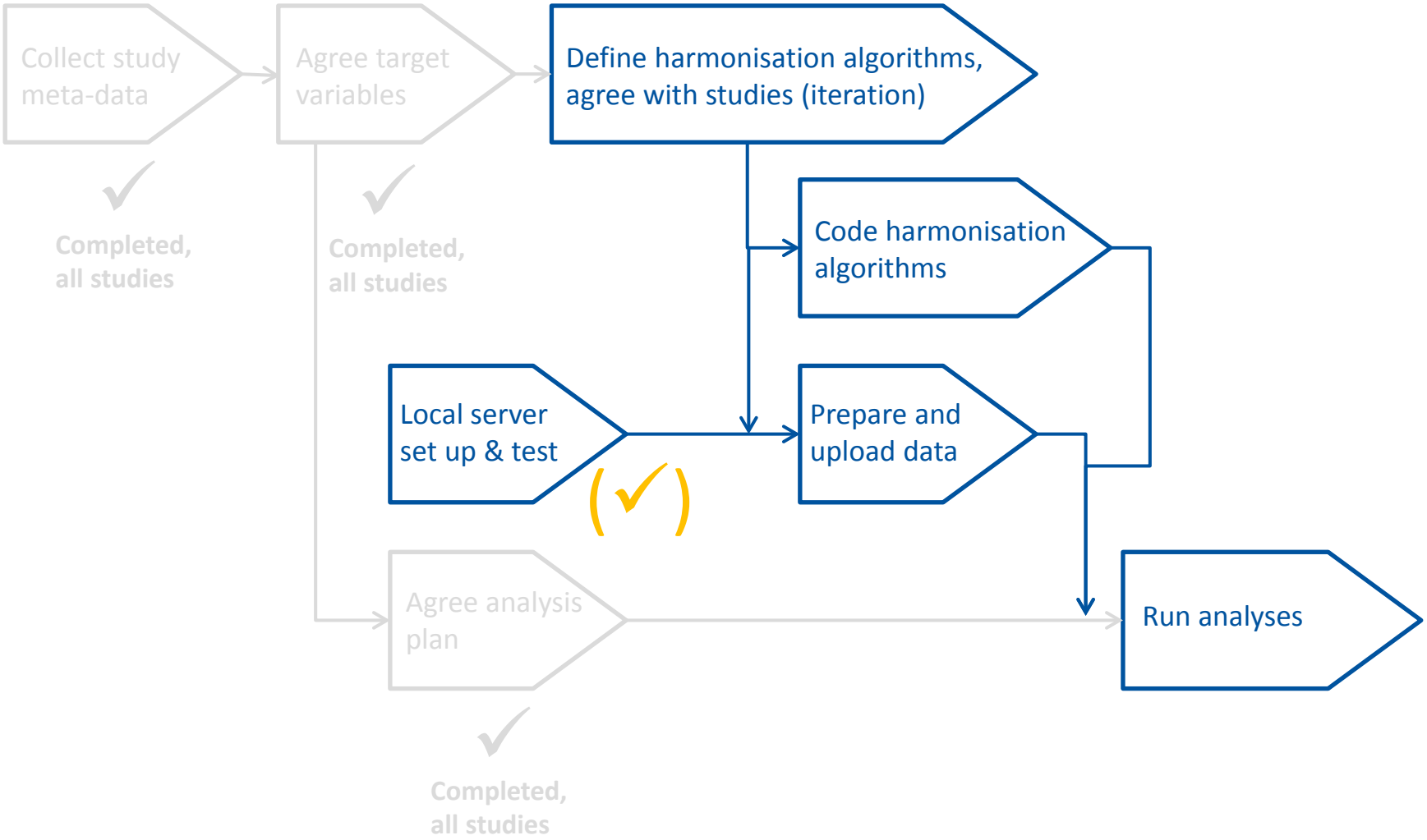
Progress


Collect study meta-data
Agree target variables
Define harmonisation algorithms, agree with studies (iteration)
Completed, all studies
Completed, all studies
Code harmonisation algorithms
Local server set up & test
Prepare and upload data
(✓)
Agree analysis plan
Run analyses
Completed, all studies

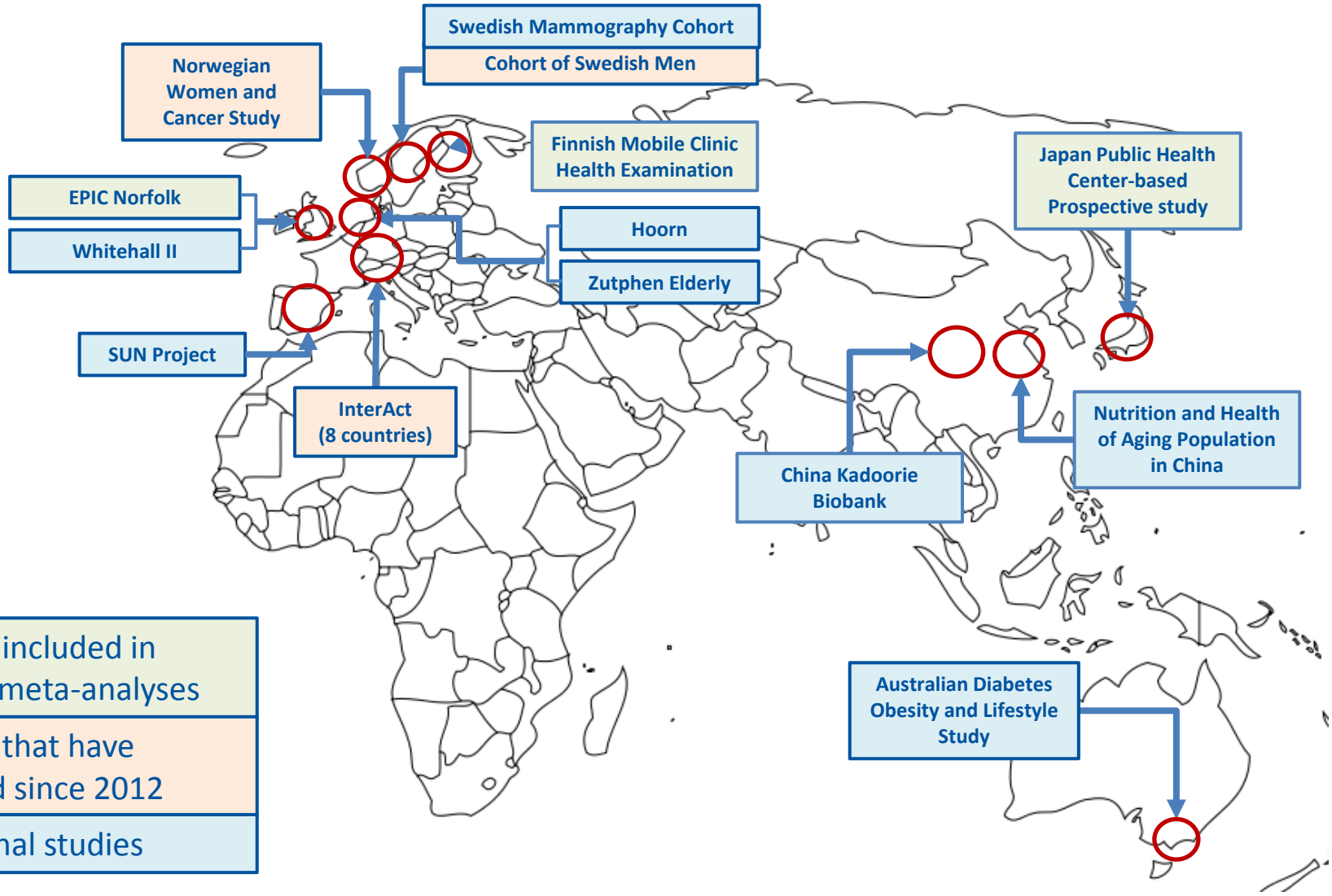
IT set up progressing

Status	Study
Completed	JPHC, NOWAC, Whitehall II
Close to completion	Hoorn, Nutrition & Health of Aging Population in China
In progress	AusDiab, EPIC-Norfolk, EPIC-InterAct, FMC, SUN, Zutphen Elderly
In discussion	SMC, COSM, China Kadoorie Biobank

Next steps


Collect study meta-data
Completed, all studies
Agree target variables
Completed, all studies
Define harmonisation algorithms, agree with studies (iteration)
Code harmonisation algorithms
Local server set up & test
(✓)
Prepare and upload data
Agree analysis plan
Completed, all studies
Run analyses

Breakdown of participating studies by previous publication on fish and T2D



3 studies included in previous meta-analyses

3 studies that have published since 2012

8 additional studies

Programme

14:30	**The InterConnect Project**	
14:45	**Applying the InterConnect approach for federated meta-analysis** 1. Physical activity in pregnancy & neonatal anthropometric outcomes 2. Fish intake and risk of type 2 diabetes	
16.00	**Future perspectives** 3. Ideas for future research projects 4. Vision and place for federated meta-analysis	 Matthias Schulze & Nita Forouhi Nick Wareham
16:30	**Discussion and involvement**	

InterConnect

Global data for diabetes and obesity research

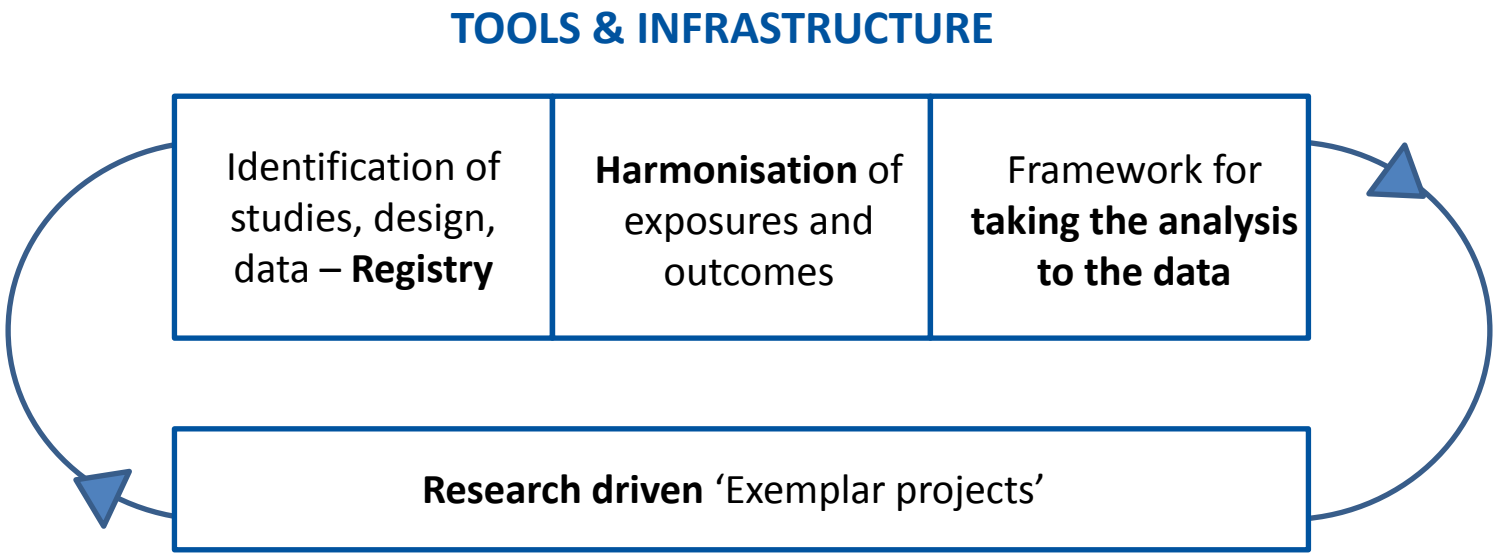
Ideas for future research projects

Matthias Schulze
German Institute of Human Nutrition, Potsdam

Nita Forouhi
MRC Epidemiology Unit, Cambridge

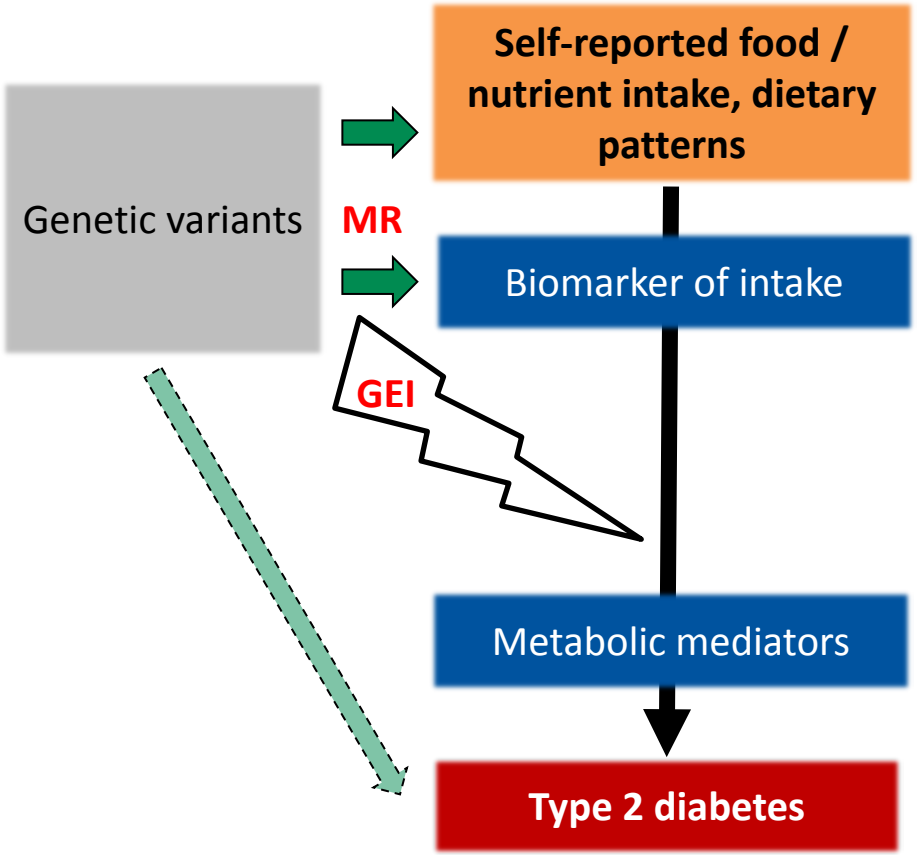
This project is funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 602068.

Developing through exemplar research projects



Potential of InterConnect for studies on diet and diabetes risk

- General framework



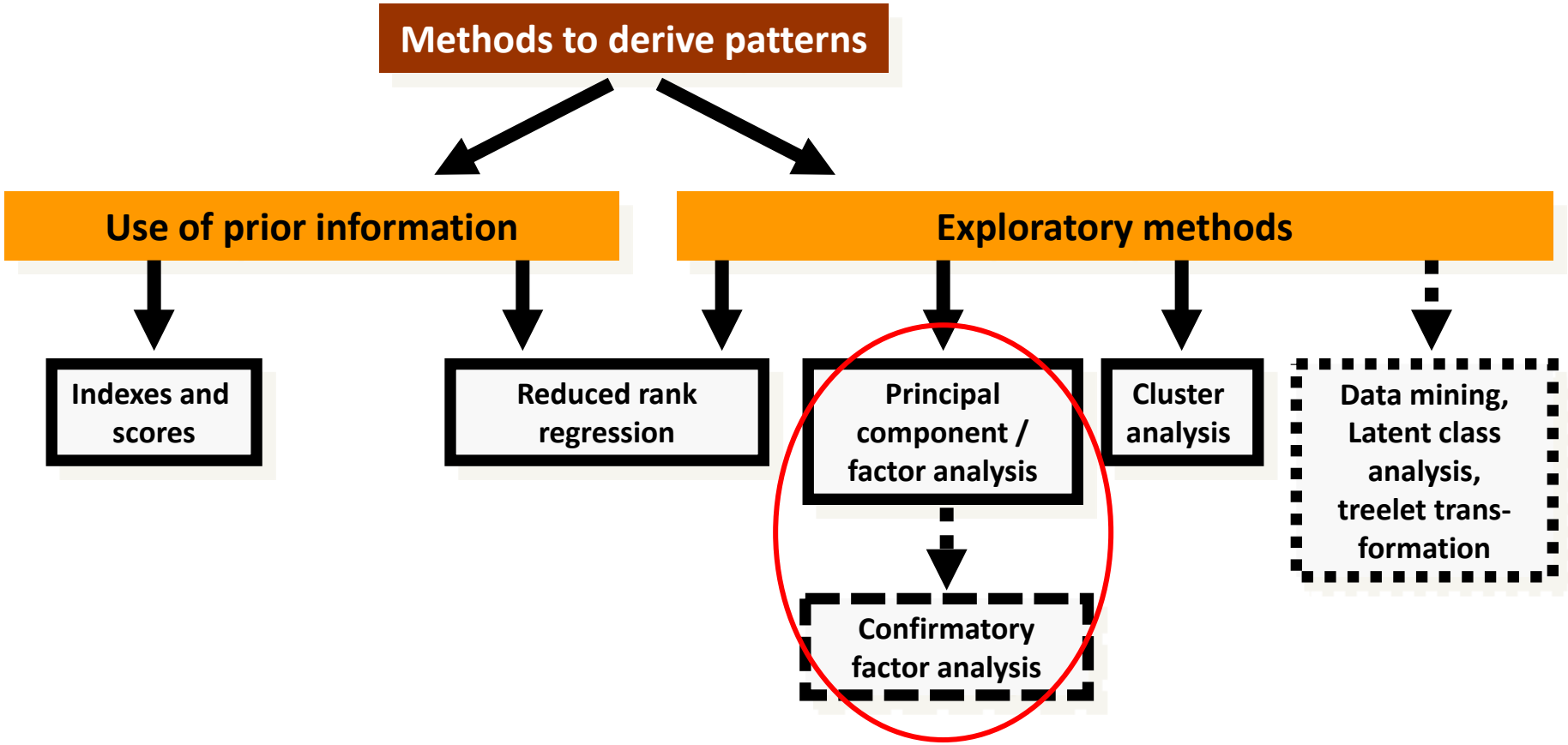
– Project ideas prioritized by InterConnect consortium members:

1. Exploratory dietary patterns
2. Gene-diet Interaction: TCF7L2
3. Legume consumption

Project 1: Exploratory dietary patterns and diabetes

Project 1: Exploratory dietary patterns and diabetes

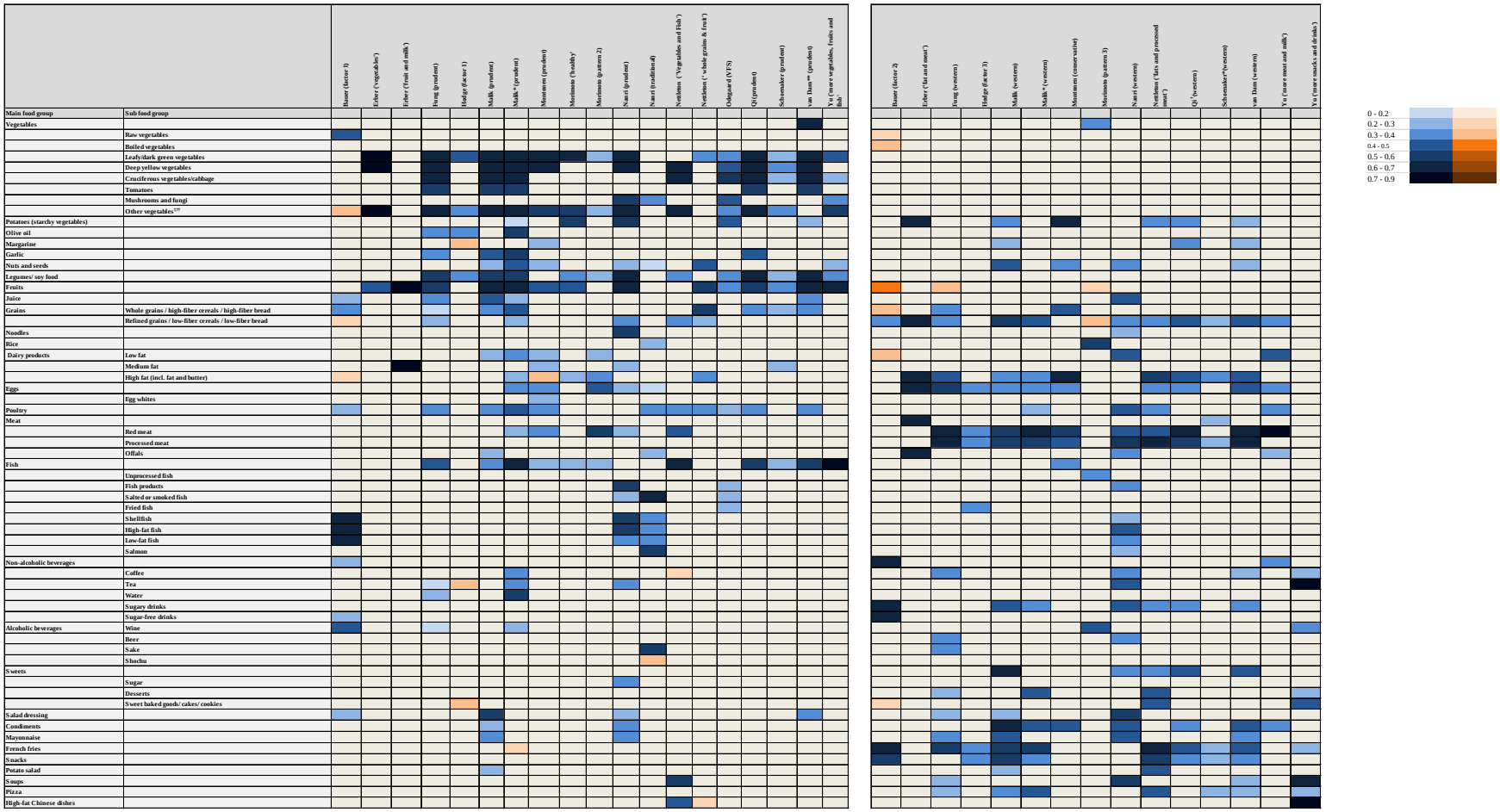
How do we identify food consumption patterns?



(adapted from Hoffmann & Schulze BJN 2007)

Exploratory dietary patterns and diabetes

Previous cohort studies on PCA/EFA and diabetes



From studies to evidence-based guidelines

Process to develop dietary guidelines

Formulate question

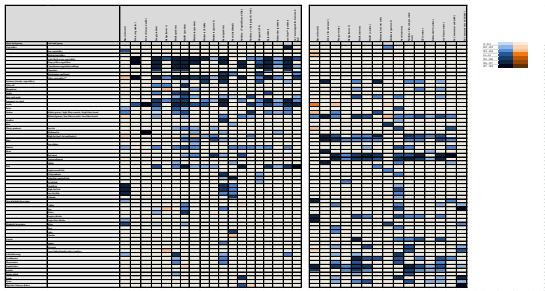
Identify literature

Extract evidence from studies

Summarize and synthesize evidence

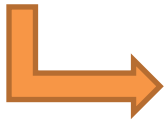
Develop and grade conclusion

How can we summarize evidence from population-specific patterns?



Project 1: Exploratory dietary patterns and diabetes

- **Aim of project**: cross-validation of previously described exploratory patterns across different populations

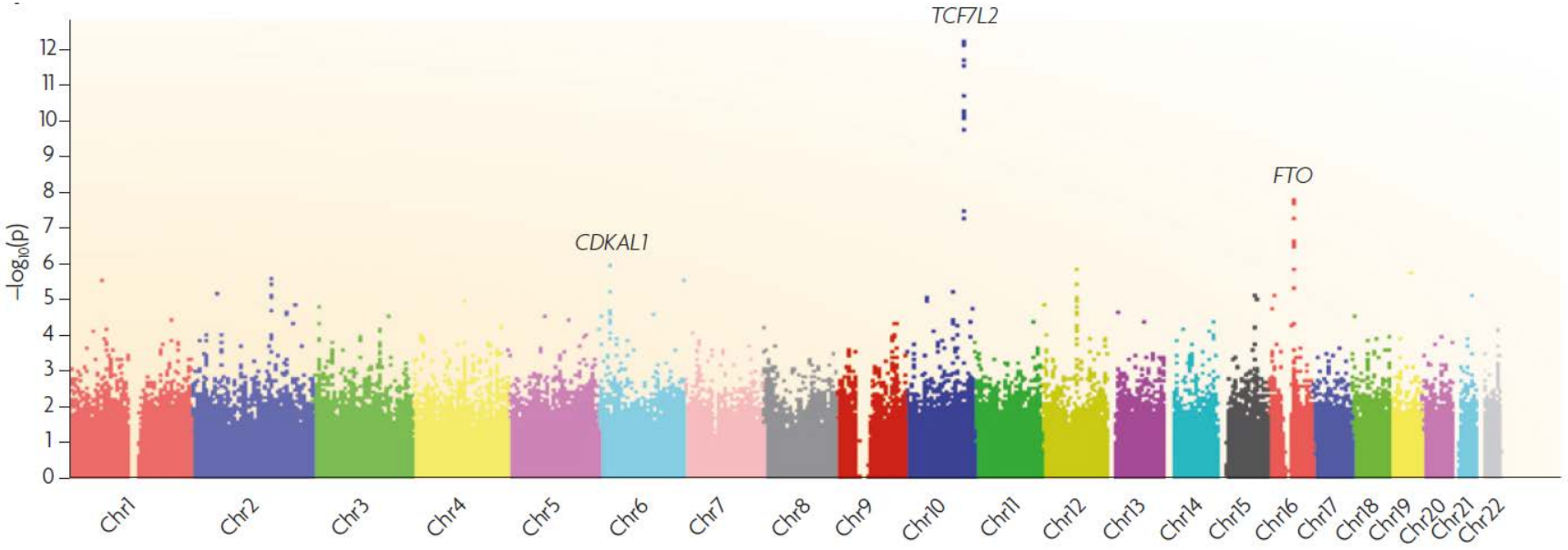


strengthen evidence-base for overall dietary patterns in diabetes prevention

Project 2: Gen-diet interactions – TCF7L2

Project 2: Gen-diet interactions – TCF7L2

- The strongest, common type 2 diabetes risk variant is located in the TCF7L2 gene

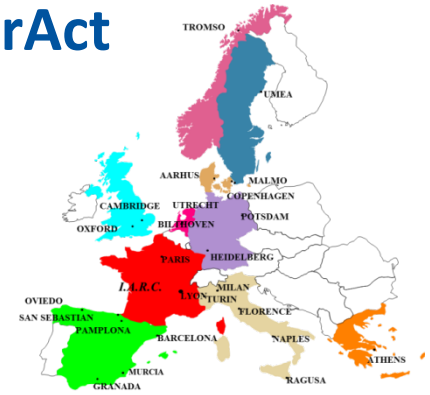


(McCarthy Nat Rev Genet 2008)

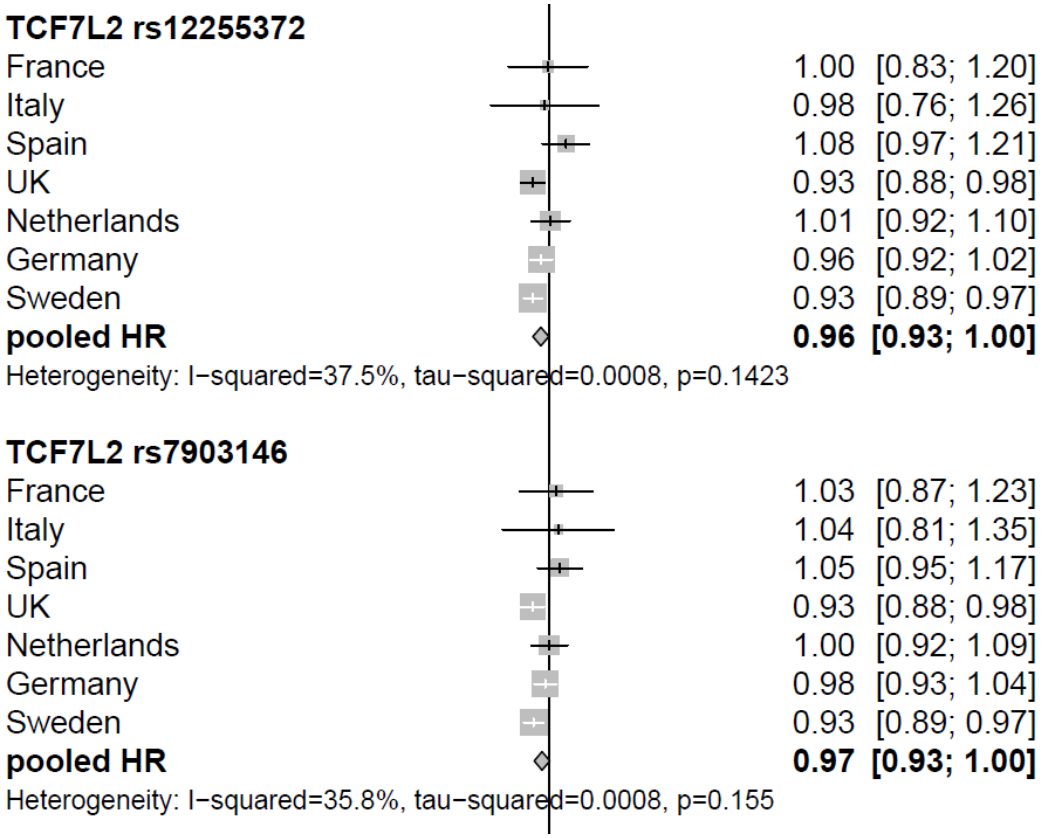
→ Can risk be modified by lifestyle?

Interaction of coffee intake and *TCF7L2* on risk of T2D

EPIC-InterAct



- ~12,000 incident diabetes cases, ~16.800 random sub-cohort
- Baseline questionnaires
- Candidate and genome-wide genotyping
- Biomarkers



(EPIC-InterAct, Diabetologia *in press*)

TcF7L2 - Previous interaction studies on T2D

lifestyle factor	interaction	no interaction
whole grain, cereal fibre, dietary fibre	Fisher (2009), Hindy (2012)	Cornelis 2009, EPIC-InterAct (in press)
GL, GI	Cornelis 2009	
dietary carbohydrate		Hindy (2012)
dietary protein		Hindy (2012)
dietary fat		Hindy (2012)
BMI		Hindy (2012), Wang (2013)
physical activity		Hindy (2012), Wang (2013)
MedDiet		Corella (2013), Langenberg (2014)

- results of previous studies not always consistent
- most studies probably underpowered.

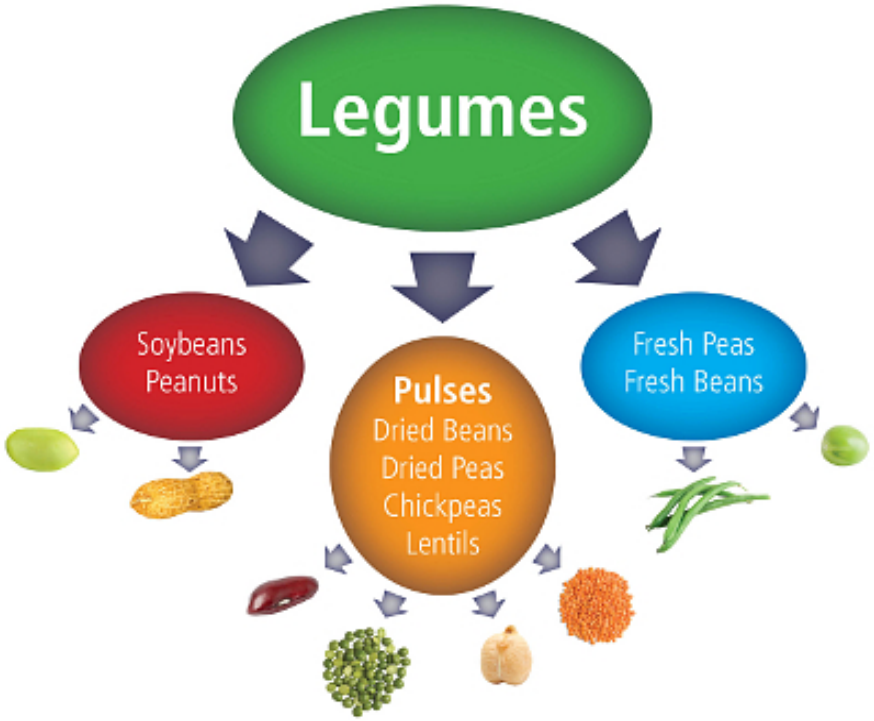
Project 2: Gene-diet Interaction - TCF7L2

Aim of project:

- De-novo meta-analysis of gene-diet interaction on risk of T2D of *TCF7L2* gene variants and the intake of
  - dietary fibre (whole grain, cereal fibre),
  - coffee
  - macronutrients on risk of T2D

Project 3: Legume intake and T2D

Project 3: Legume intake and T2D



- **Why should we be interested?**
- Legumes are excellent sources of protein, dietary fibre, and a variety of micronutrients and phytochemicals, & low in fat
- In line with WHO recommendation to limit the consumption of red and processed meat, legumes could provide healthy alternatives to meat products
- Amount and type of legume consumption varies substantially across the world, & people from many regions consume legumes as staple foods, but health effects are largely unknown

What is the research evidence?

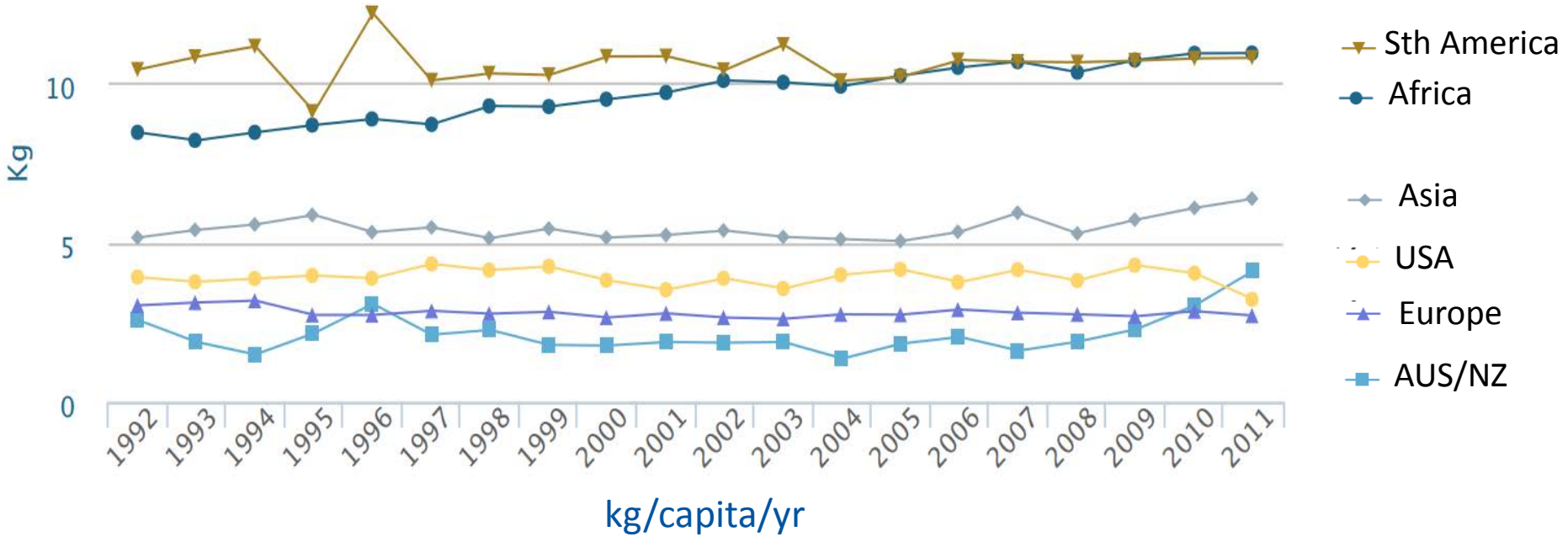
Prospective cohort studies reporting measures of association between intake of legumes in relation to type 2 diabetes

Authors	Date	Population	N of T2D cases	Legume intake	Exposure measure	Association	P_{trend}
A. M. Hodge *et al.*	2004	MCCS (AUSTRALIA)	365	Total (g/d)	FFQ	OR= 1.09 (0.81-1.47)	0.7
S. Liu *et al.*	2004	WHS (USA)	1,608	Total (sv/d)	FFQ	RR= 1.12 (0.95-1.33)	0.2
L. A. Bazzano *et al.*	2008	NHS (USA)	4,529	Total (sv/d)	FFQ	HR = 1.14 (1.03-1.25)	0.09
R.Villegas *et al.*	2008	SWHS (ASIA)	1,605	Total (g/d)	FFQ	**RR= 0.62 (0.51-0.74)**	<0.001
M. Aldwairji *et al.*	2013	UK WCS (EUROPE)	114	Total (g/d)	FFQ	OR = 1.33 (0.80-2.22)	0.25
				Dried (g/d)	FFQ	**OR = 0.85 (0.52-0.84)**	0.03
				Fresh (g/d)	FFQ	OR = 1.51 (0.89-2.59)	0.27

Conflicting for total legumes, and unavailable for types of legumes

Variation of types of legume intake across the world

Pulses

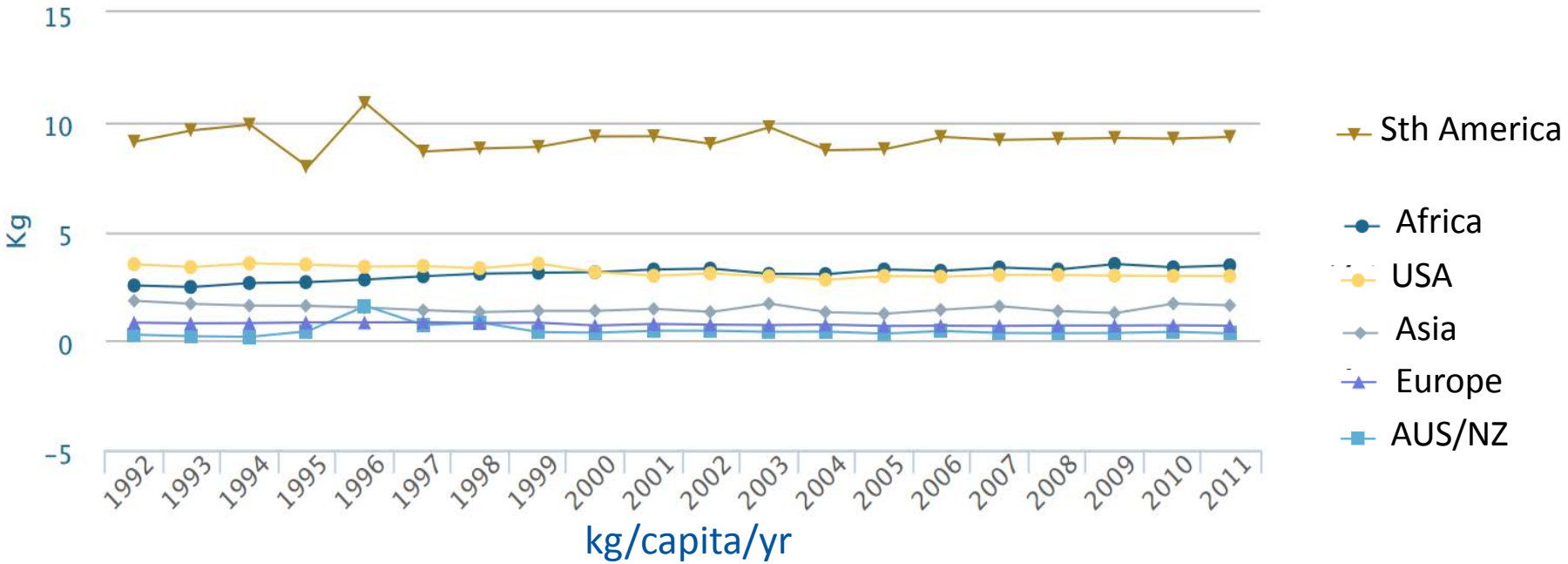


kg/capita/yr

FAOSTAT

Variation of types of legume intake across the world

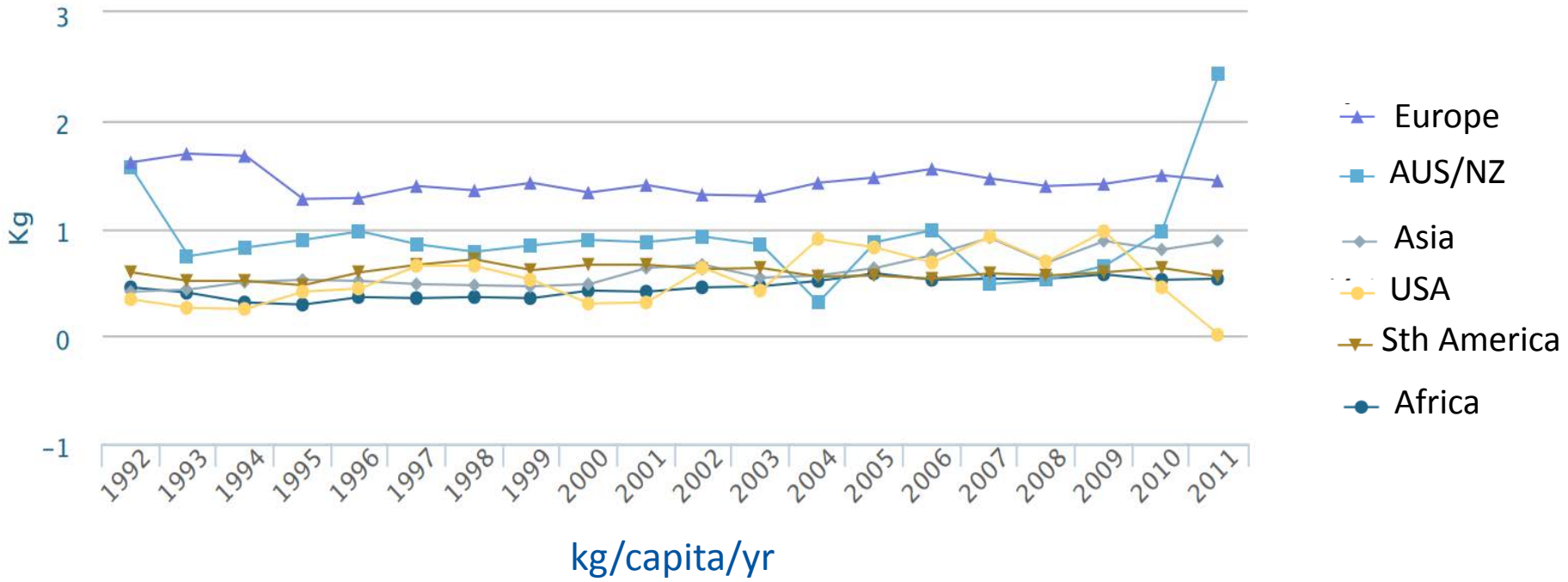
Beans



FAOSTAT

Variation of types of legume intake across the world

Peas

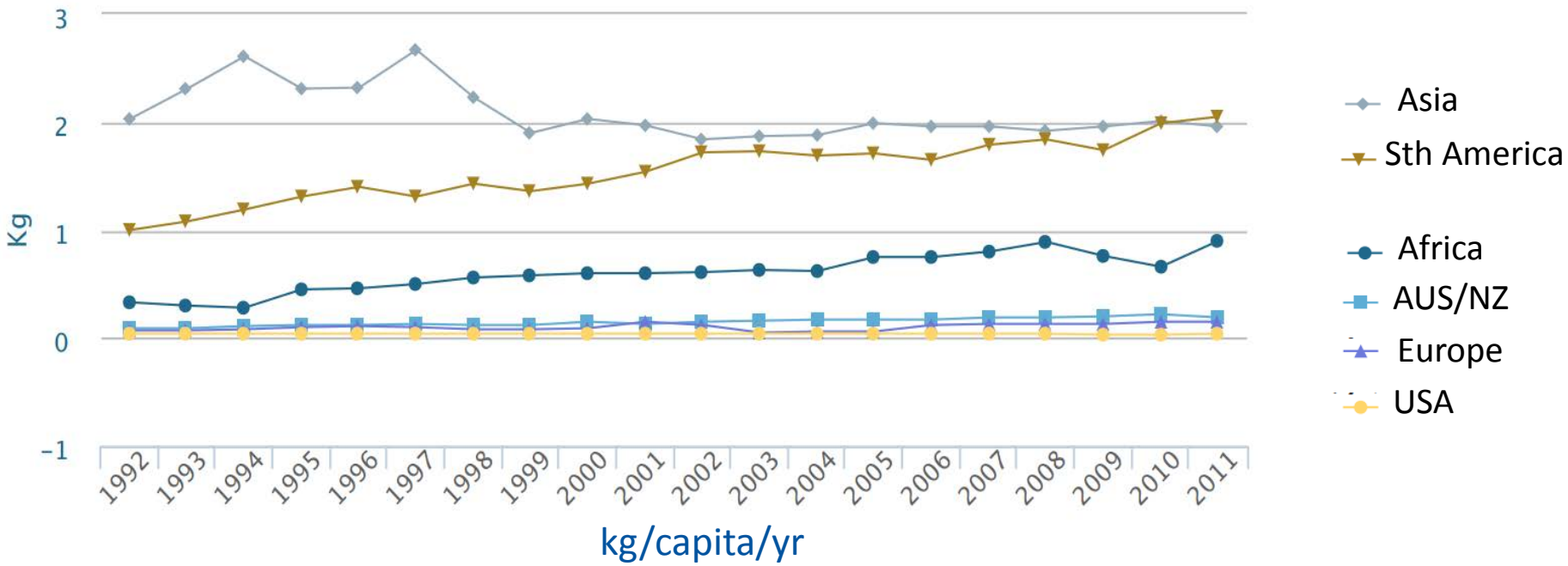


kg/capita/yr

FAOSTAT

Variation of types of legume intake across the world

Soyabeans



FAOSTAT

Legume intake and T2D – Why InterConnect is the right approach to investigate

- Public health importance of the research question
- Limited available evidence
- We know that some cohorts worldwide have data on legumes but have not investigated their association with T2D
- Federated analysis approach of InterConnect offers the opportunity to
  - standardise the definitions of types of legumes,
  - consider the geographical variation in legume intake,
  - account for confounding factors using individual participant level data,
  - without data leaving research institutions

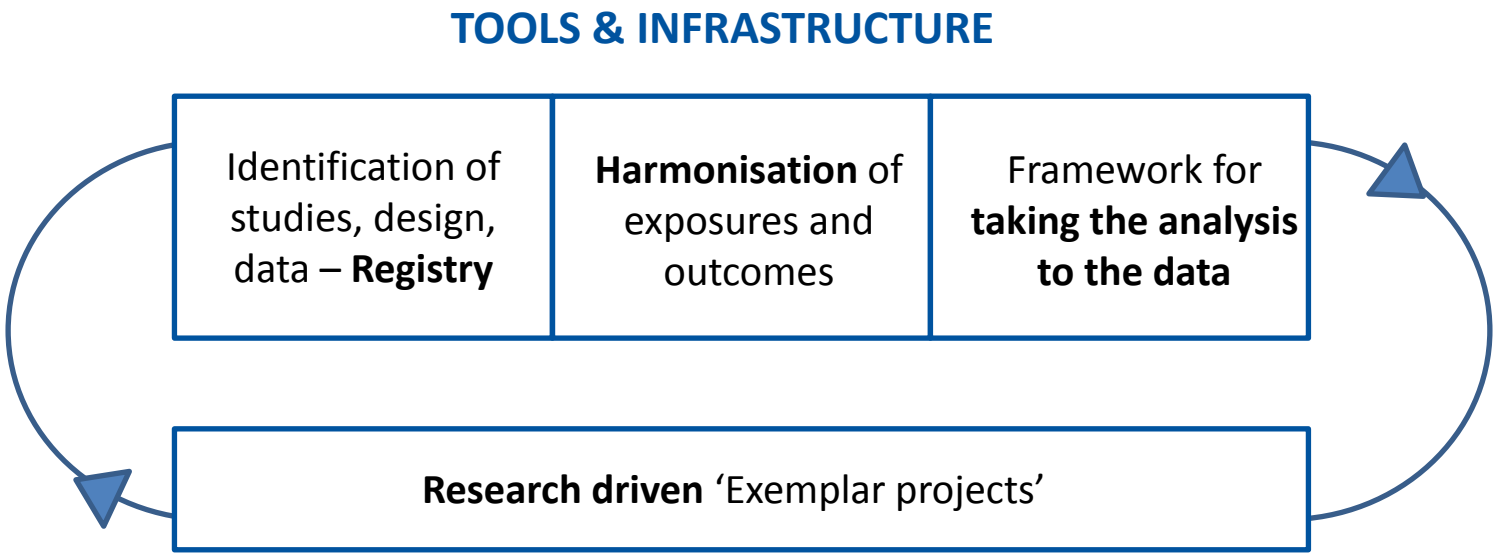
Setup of 3 new InterConnect projects

- General criteria for participating studies:
  - **Population:** Include general population; no limits on age; exclude prevalent cases of type 2 diabetes
  - **Exposure:**
    - food and beverage consumption (amount/period and/or frequency of intake) and derived intake of dietary fibre and macronutrients
    - TCF7L2 gene polymorphisms (for project 2 only)
  - **Outcome:** incidence of T2D: self-reported or objectively measured (information on diagnosis date would be ideal)
  - **Potential Confounders:** Age, sex, smoking, body mass index (BMI), physical activity, socio-economic status (education or occupation), family history of diabetes, other health exposures (cardiovascular diseases, history of previous illness)
  - **Modifiers:** Geographic location

It’s not only about diet

- A 4th proposed exemplar is on childhood growth and development
- The association between birthweight and later adiposity – to be led by Dr Ken Ong and Dr Gernot Desoye

Developing through exemplar research projects



InterConnect

Global data for diabetes and obesity research

Vision and place for InterConnect approach to federated meta-analysis

Nick Wareham

12 September 2016

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

InterConnect approach

- PA in pregnancy exemplar project
  - Analysis equivalent to a meta-analysis of harmonised individual level participant data (IPD)
  - Delivering scientifically interesting results of public health relevance
- No physical sharing or pooling data
- Achieved via access to data and sharing of results
  - Access is ‘blind’ – individual records are neither visible nor physically accessible

Overview of process – some new tasks

Tasks specific to InterConnect approach are highlighted in italics

Others are common research tasks

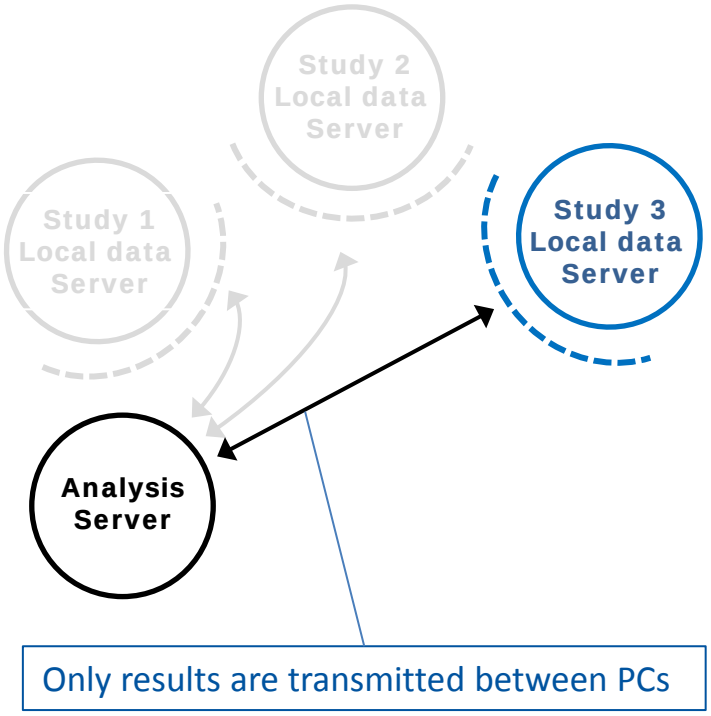
ROLES	TASKS				
Coordinator	Define question	Identify studies	Agree analysis plan	(→ on-going coordination)	
Studies	Provide meta-data	*Set up local server*	Review harmonisation algorithms	*Upload data*	
Harmonisation Lead	Review meta-data	Define target variables	Propose harmonisation algorithms		
Technical Lead	*Network set up & security (→ on-going)*	*Support study server set up*	*Check server connectivity*	*Code and run harmonisation algorithms*	*Run analyses (R, DataSHIELD*

Place of InterConnect approach

- Some new tasks - cost and time to set up
  - Cost of server – dependent on local context and study size
  - Time to set up server (c. 12 – 16 hours), uploading dataset (c. 8 hours), plus on-going patches and software upgrades
- Use only where needed
  - Contextual variation - environmental or place based differences in risk - and /or genetic differences in risk
  - Can’t be addressed within a single study

Benefits once set up

- Study effort focused on IT and preparation of data
  - Re-use by giving access to additional sub-sets of data to address new questions
  - Studies remain in control of data – unlike central deposition
- Unlike traditional meta-analysis
  - Studies don't need to perform any analyses – work follows role
  - Analysis is done in real time – no wait for outputs from study analysts

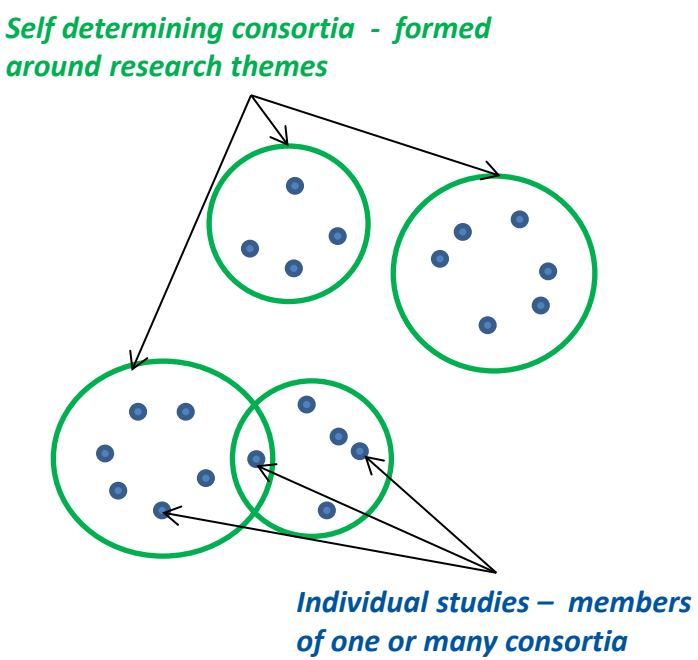


Vision

- A global ‘data access and results sharing’ network for federated meta-analysis of harmonised individual level participant data

Autonomous consortia within network

- InterConnect is NOT an analytical consortium
- Seeks to enable others
  - ad hoc consortia form within a framework
  - each to decide its own way of working and be autonomous



Delivering the vision – through exemplars

- Broadening studies
- Broadening roles
- Enabling management processes

Broadening studies: PA (8) & fish (13) exemplars

Europe	UK	Sweden	Finland	China
EPIC-InterAct	ALSPAC	Mammography	Mobile Clinic Health	Nutrition & Health
Netherlands	SWS	Swedish Men	**Norway**	Kadorie Biobank
ABCD	Whitehall II Study	**Poland**	NOWAC	**Japan**
Gecko Drenthe	**Spain**	Repro_PL	**Ireland**	JPHC
Hoorn Study	Sun Project	**Denmark**	ROLO	**Australia**
Zutphen Elderly		DNBC	**USA**	AusDiab
			Healthy Start Study	

Broadening roles

ROLES	TASKS				
Coordinator	Define question	Identify studies	(→ on-going coordination)		
Studies	Provide meta-data	Set up local server	Review harmonisation algorithms	Upload data	
Harmonisation Lead	Review meta-data	Define target variables	Propose harmonisation algorithms		
Technical Lead	Network set up & security (→ on-going)	Support study server set up	Check server connectivity	Code and run harmonisation algorithms	Run analyses (R, DataSHIELD

InterConnect has been leading the exemplars

Broadening roles: others can lead

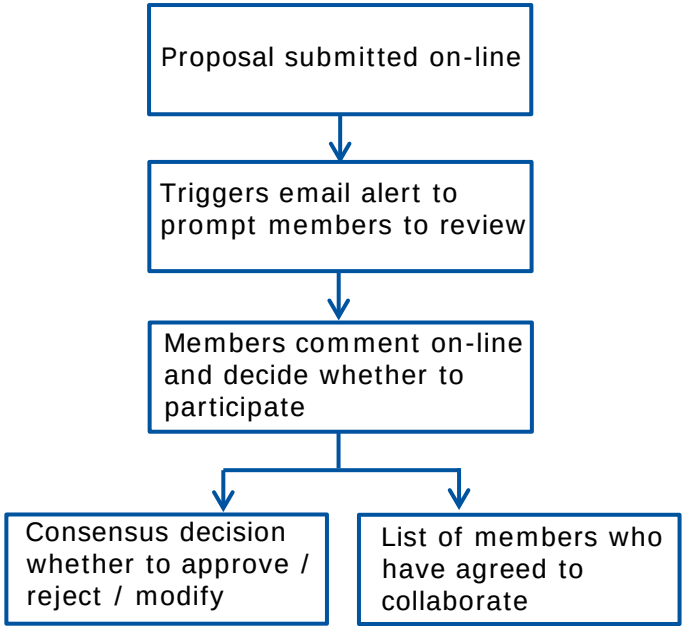
ROLES	TASKS				
Coordinator	Define question	Identify studies	(→ on-going coordination)		
Studies	Provide meta-data	Set up local server	Review harmonisation algorithms	Upload data	
Harmonisation Lead	Review meta-data	Define target variables	Propose harmonisation algorithms		
Technical Lead	*Network set up & security (→ on-going)*	*Support study server set up*	*Check server connectivity*	*Code and run harmonisation algorithms*	*Run analyses (R, DataSHIELD*

Others can lead

InterConnect continues technical support

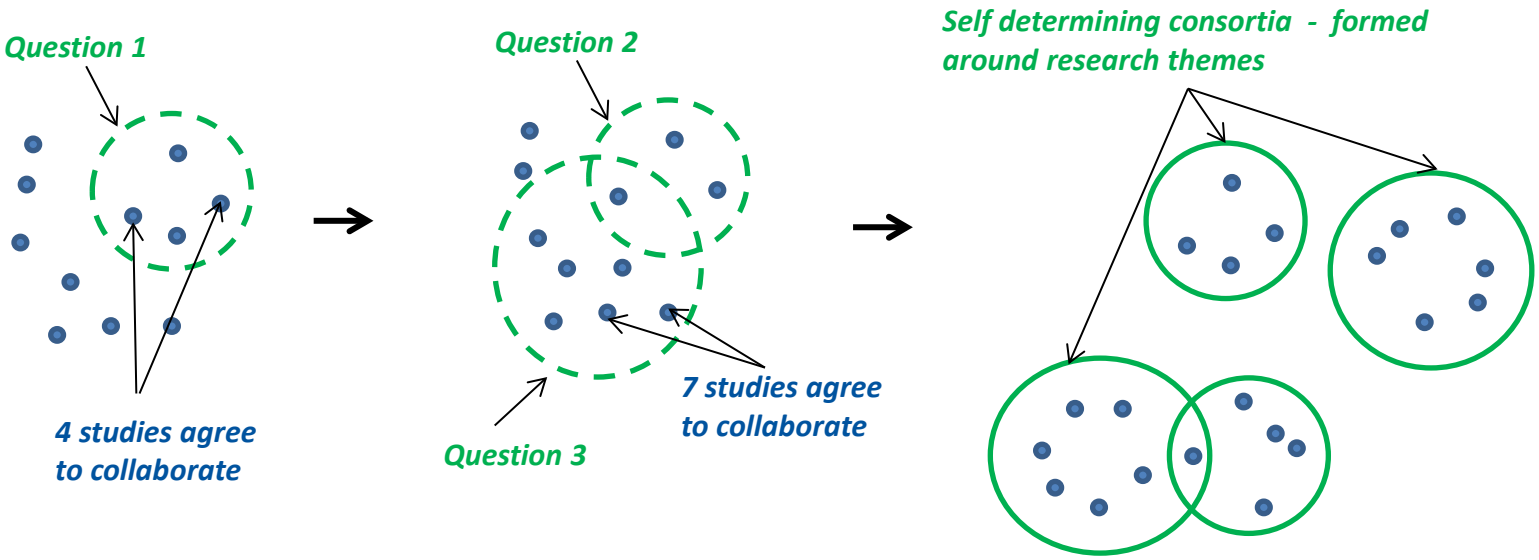
Enabling management processes

- ‘Data access and results sharing’ collaboration agreement
- Transparent and democratic processes through online tool – in development



Vision: step-wise

- Initially studies coalesce / disperse around exemplar questions
- Research themes and more stable groupings emerge
- Initially a facilitated network → more autonomous consortia



Getting involved (1) Registry

Short Name	Name	Study Design	Actual number of participants recruited to the study	Country of residence
MEC	Multiethnic Cohort Study	Prospective cohort study	215 251	United States
SWS	Southampton Women's Survey	Prospective cohort study	12 583	United Kingdom
	Healthy Start study	Prospective cohort study	2 820	United States
ALSPAC	Avon Longitudinal Study			
AHS	Agricultural Health Study			
ARIC	Atherosclerosis Risk in C			
DNBC	Danish National Birth Co			
EPIC - Turin	European Prospective In			
NHS I	Nurses Health Study I			
NOMAS	The NOrthern MAnhattar			

Our approach is to enable wide coverage of studies with a limited set of information that can largely be collated from information already in the public domain. This creates little burden for each individual study while enabling sign-posting of a large number of studies useful for cross-cohort analyses. Meta-data currently included in the registry comprises:

- Study name, contact, reference paper, website
- Study design, timeline, number of participants
- Broad categories of ethnic and racial groups recruited
- Health information at baseline and during follow up, as well as key exposures
- Participant selection criteria and recruitment procedures

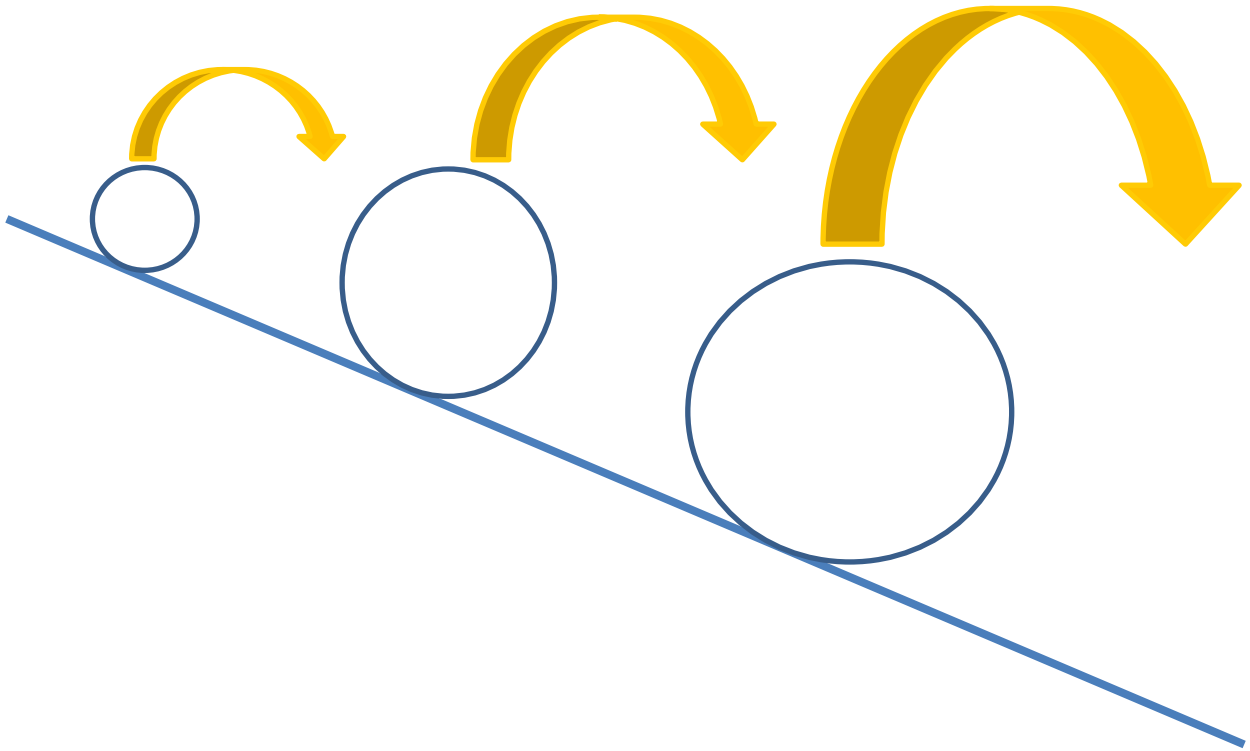
You can view and search the registry here.

Email InterConnectRegistry@mrc-epid.cam.ac.uk if you would like to include your study in the registry

Getting involved (2) New exemplars

- Exploratory dietary patterns
- Gene-diet Interaction: TCF7L2
- Legume consumption
- Birthweight and later adiposity
- Others?

Together we can create a snowball



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