

# 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

Received: 10 December 2010 / Accepted: 4 April 2011  
© Springer-Verlag 2011

## Abstract

**Aims/hypothesis** Studying gene–lifestyle interaction may help to identify lifestyle factors that modify genetic susceptibility and uncover genetic loci exerting important subgroup effects. Adequately powered studies with prospective, unbiased, standardised assessment of key behavioural factors for gene–lifestyle studies are lacking. This case–cohort study aims to investigate how genetic and potentially modifiable lifestyle and behavioural factors, particularly diet and physical activity, interact in their influence on the risk of developing type 2 diabetes.

**Methods** Incident cases of type 2 diabetes occurring in European Prospective Investigation into Cancer and Nutrition (EPIC) cohorts between 1991 and 2007 from eight of the ten EPIC countries were ascertained and verified. Prentice-weighted Cox regression and random-effects meta-analyses were used to investigate differences in diabetes incidence by age and sex.

**Results** A total of 12,403 verified incident cases of type 2 diabetes occurred during 3.99 million person-years of follow-up of 340,234 EPIC participants eligible for Inter-

Act. We defined a centre-stratified subcohort of 16,154 individuals for comparative analyses. Individuals with incident diabetes who were randomly selected into the subcohort ( $n=778$ ) were included as cases in the analyses. All prevalent diabetes cases were excluded from the study. InterAct cases were followed-up for an average of 6.9 years; 49.7% were men. Mean baseline age and age at diagnosis were 55.6 and 62.5 years, mean BMI and waist circumference values were 29.4 kg/m<sup>2</sup> and 102.7 cm in men, and 30.1 kg/m<sup>2</sup> and 92.8 cm in women, respectively. Risk of type 2 diabetes increased linearly with age, with an overall HR of 1.56 (95% CI 1.48–1.64) for a 10 year age difference, adjusted for sex. A male excess in the risk of incident diabetes was consistently observed across all countries, with a pooled HR of 1.51 (95% CI 1.39–1.64), adjusted for age.

**Conclusions/interpretation** InterAct is a large, well-powered, prospective study that will inform our understanding of the interplay between genes and lifestyle factors on the risk of type 2 diabetes development.

**Keywords** Cohort studies · Diabetes mellitus · Epidemiologic research design · Gene–lifestyle interaction · Genetic epidemiology · Type 2 diabetes

The InterAct Consortium list of authors is shown in the [Appendix](#).

**Electronic supplementary material** The online version of this article (doi:10.1007/s00125-011-2182-9) contains supplementary material, which is available to authorised users.

C. Langenberg (✉), Medical Research Council Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Box 285, Cambridge CB2 0QQ, UK  
e-mail: claudia.langenberg@mrc-epid.cam.ac.uk

The InterAct Project URL: <http://www.inter-act.eu>

## Abbreviations

EPIC	European Prospective Investigation into Cancer and Nutrition
GEI	Gene–environment or lifestyle interaction
GWAS	Genome-wide association studies
IARC	International Agency for Research into Cancer
MRC	Medical Research Council
SNP	Single nucleotide polymorphism

## Introduction

The identification of common genetic variants that are reproducibly associated with type 2 diabetes has accelerated considerably with the availability of results from genome-wide association studies (GWAS) of prevalent cases and controls [1–6]. However, the associations discovered only account for a relatively small proportion of the heritable component [7]. Interactions between genetic factors and lifestyle exposures, gene–gene interaction and genetic variation other than common single nucleotide polymorphisms (SNPs) are likely to be important factors that contribute to the remaining variance [8], but have not been systematically explored.

The existing case–control studies used to identify the genetic loci associated with type 2 diabetes are not optimally designed to investigate gene–environment or lifestyle interaction (GEI) as they do not have standardised assessment of key lifestyle or behavioural factors and do not have a prospective design in which those factors are assessed in an unbiased manner before the onset of disease. An optimal study design for investigating GEI is a case–cohort study nested within a large prospective cohort, as this combines the efficiency of the case–control design with the advantages of the longitudinal cohort approach with extensive prospective assessment of key exposures that are not subject to recall bias. Selecting a random subcohort (nested case–cohort study) rather than matched controls (nested case–control study) has the additional advantage that it facilitates the design and conduct of future case–cohort studies for other diseases occurring in the same background population or cohort.

The InterAct Consortium is funded by the Sixth European Framework Programme. It was initiated to investigate how genetic and lifestyle behavioural factors, particularly diet and physical activity, interact on the risk of developing diabetes and how knowledge about such interactions may be translated into preventive action. As part of the wider InterAct Project, consortium partners have established a case–cohort study of incident type 2 diabetes (InterAct Study) based on cases occurring in European Prospective Investigation into Cancer and Nutrition (EPIC) cohorts between 1991 and 2007 in eight of ten EPIC countries participating in InterAct.

The principal objectives of this report are to: (1) describe the InterAct Study design, population and objectives; (2) characterise the random subcohort and compare it with EPIC participants from each of the eight participating European countries eligible for InterAct; and (3) investigate characteristics of incident diabetes cases, including differences in diabetes risk by age and sex.

## Methods

*Participants and study design* The large prospective InterAct type 2 diabetes case–cohort study is coordinated by the Medical Research Council (MRC) Epidemiology Unit in Cambridge and nested within EPIC [9]. EPIC was initiated in the late 1980s and involves collaboration between 23 research institutions across Europe in ten countries (Denmark, France, Germany, Greece, Italy, the Netherlands, Norway, Spain, Sweden and the UK). With the exception of Norway and Greece, all EPIC countries participated in the InterAct Project, with a total of 455,680 participants (Table 1). The majority of EPIC cohorts were recruited from the general population, with some exceptions [10]. French cohorts included women who were members of a health insurance scheme for school and university employees; Turin and Ragusa (Italy) and the Spanish centres included some blood donors. Participants from Utrecht (the Netherlands) and Florence (Italy) were recruited via a breast cancer screening programme. The majority of participants recruited by the EPIC Oxford (UK) centre were vegetarian and ‘health conscious’ volunteers from England, Wales, Scotland and Northern Ireland [10].

All participants gave written informed consent, and the study was approved by the local ethics committee in the participating countries and the Internal Review Board of the International Agency for Research on Cancer.

*Measurements* As part of EPIC, standardised information was collected at baseline on lifestyle exposures. Information on socioeconomic status, education and occupation was collected by questionnaire.

The assessment of diet was undertaken using a self- or interviewer-administered dietary questionnaire, developed and validated within each country to estimate the usual individual food intakes of the study participants [10, 11]. Additionally, in a stratified subsample of 36,900 participants, a standardised 24 h recall of food intake was collected [11, 12].

Physical activity was assessed at baseline using a brief questionnaire covering occupation and recreational activity [12, 13]. Although an index of physical activity derived from this questionnaire had previously been validated against repeated objective measures of activity in the UK and in the Netherlands [13], no validation study has been conducted in any other country. Therefore, a study to test the validity of the questionnaire in European populations was conducted in InterAct, including 100 men and 100 women comparable with those originally recruited into the EPIC study from each participating InterAct country, with an objective measurement of physical activity by individually calibrated combined heart rate and movement sensing [14]. Results from this validation study will be published as a separate report.

**Table 1** Overview of the EPIC cohorts contributing to the InterAct type 2 diabetes case-cohort study

Country	Centre	Baseline collection		Stored samples <i>n</i>	Women (%)	5th and 95th age percentiles
		Period	<i>n</i>			
France	Ile-de-France	1993–1996	14,196	5,202	100	44–65
	North-west France	1993–1996	13,073	3,880	100	44–65
	North-east France	1993–1996	16,244	3,919	100	44–65
	Rhone-Alpes/Auvergne	1993–1996	9,966	3,319	100	44–65
	Provence/Languedoc	1993–1997	10,126	2,690	100	44–65
	South-west France	1993–1996	10,919	2,076	100	44–65
Italy	Florence	1992–1998	13,597	13,597	74	37–63
	Varese	1993–1997	12,083	12,073	79	39–64
	Ragusa	1993–1997	6,403	6,397	52	36–61
	Turin	1993–1998	10,604	10,604	43	37–62
	Naples	1993–1997	5,062	5,057	100	38–63
Spain	Asturias	1992–1995	8,542	8,422	64	36–62
	Granada	1992–1996	7,879	6,892	77	36–64
	Murcia	1992–1996	8,516	8,146	68	36–62
	Navarra	1992–1995	8,084	8,030	52	37–62
	San Sebastian	1992–1995	8,417	8,338	51	37–62
UK	Cambridge	1993–1998	30,441	24,036	55	45–74
	Oxford	1993–1998	57,489	19,241	77	24–70
Netherlands	Bilthoven	1993–1997	22,715	19,388	55	23–58
	Utrecht	1993–1997	17,357	16,930	100	49–68
Germany	Heidelberg	1994–1998	25,540	24,236	53	37–63
	Potsdam	1994–1998	27,548	26,444	60	36–64
Sweden	Malmö	1991–1996	28,098	28,053	61	47–71
	Umeå	1992–1996	25,728	25,728	52	30–60
Denmark	Aarhus	1995–1997	17,154	17,094	51	50–64
	Copenhagen	1993–1997	39,899	39,036	53	50–64
Total			455,680	348,828 <sup>a</sup>		

<sup>a</sup> Total number of samples indicates storage of any blood sample; InterAct eligibility criteria additionally included availability of a buffy coat sample for DNA extraction

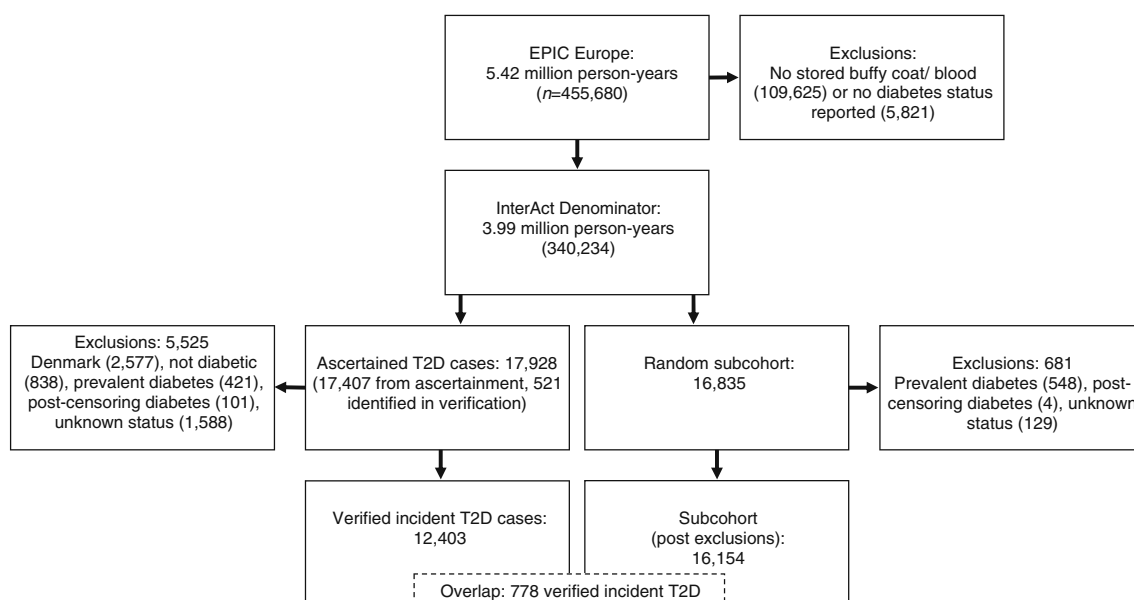
Standard anthropometric data and biological samples (blood plasma, blood serum, white blood cells and erythrocytes) were collected from 385,747 of the 519,978 EPIC study participants and 346,055 of 455,680 individuals participating in eight of the ten EPIC cohorts included in InterAct. Individuals without stored blood ( $n=109,625$ ) or without information on reported diabetes status ( $n=5,821$ ) were excluded, leaving a total of 340,234 participants with 3.99 million person-years of follow-up eligible for inclusion in InterAct (Fig. 1).

Samples were stored from collection at  $-196^{\circ}\text{C}$  in liquid nitrogen at the coordinating centre at the International Agency for Research into Cancer (IARC) in Lyon, France, or in liquid nitrogen in local biorepositories with the exception of Umeå, where  $-80^{\circ}\text{C}$  freezers were used. Follow-up data on mortality and disease status has been ascertained via registries, clinical records and other sources

of clinical information [15, 16]. At least one follow-up was conducted in each centre 3–5 years after baseline and questionnaires and telephone-based interviews were administered to repeat exposure measurement and collect self-reported health status data.

*Type 2 diabetes case ascertainment and verification* We followed a pragmatic high-sensitivity approach for case ascertainment with the aim of identifying: (1) all potential incident diabetes cases; and (2) excluding all individuals with prevalent diabetes.

Prevalent diabetes was identified on the basis of baseline self-report of a history of diabetes, doctor-diagnosed diabetes, diabetes drug use or evidence of diabetes after baseline with a date of diagnosis earlier than the baseline recruitment date. All ascertained cases with any evidence of diabetes at baseline were excluded.



**Fig. 1** Overview of the InterAct type 2 diabetes (T2D) case-cohort study nested within eight of the ten EPIC Europe countries

Ascertainment of incident type 2 diabetes involved a review of the existing EPIC datasets at each centre using multiple sources of evidence including self-report, linkage to primary-care registers, secondary-care registers, medication use (drug registers), hospital admissions and mortality data (electronic supplementary material [ESM] Table 1). Information from any follow-up visit or external evidence with a date later than the baseline visit was used. Cases in Denmark and Sweden were not ascertained by self-report, but identified via local and national diabetes and pharmaceutical registers and hence all ascertained cases were considered to be verified (ESM Table 1).

To increase the specificity of the case definition for centres other than those from Denmark and Sweden, we sought further evidence for all cases with information on incident type 2 diabetes from fewer than two independent sources at a minimum, including individual medical records review in some centres. Follow-up was censored at the date of diagnosis, 31 December 2007 or the date of death, whichever occurred first. In total, 12,403 verified incident cases were identified; there were 471 cases in the first year of follow-up and 587 in the second year. Sample size calculations are included in the online appendix (ESM Fig. 1).

**Dates of diagnosis** The date of diagnosis for incident cases was set as: the date of diagnosis reported by the doctor; the earliest date that diabetes was recorded in medical records; the date of inclusion into the diabetes registry; the date reported by the participant; or the date of the questionnaire in which diabetes was first reported. If the date of diagnosis could not be ascertained from any of the sources listed above, the midpoint between recruitment and censoring was used ( $n=421$ ).

**Case-cohort design** The case-cohort design of the InterAct Study differs from the nested case-control design in that a random subcohort is selected instead of a set of matched controls. Because only a subset of the original cohort is randomly selected into the subcohort, cases are over-represented in the case-cohort set, which needs to be accounted for in the analysis methods, as outlined in the ESM.

**Subcohort** A subcohort of 16,835 individuals was randomly selected from those with available stored blood and buffy coat, stratified by centre. We oversampled the number of individuals in the subcohort for the proportion of prevalent diabetes cases in each centre to account for later exclusion of individuals with prevalent diabetes from InterAct analyses. After exclusion of 548 individuals with prevalent diabetes, 129 individuals with unknown diabetes status and four individuals with post-censoring diabetes, 16,154 subcohort individuals were included in the analysis. Because of random selection, this subcohort also included a random set of 778 individuals who had developed incident type 2 diabetes during follow-up.

**Stored samples, genotyping and biomarker measurement** Details of the quality, quantity and availability of stored samples can be found in the ESM, together with a description of the InterAct strategy for genotyping and biomarker measurement.

**Statistical analyses** Characteristics of the InterAct incident cases are described using summary statistics (means, standard deviations, frequencies and percentages) separately for men and women, and overall. Characteristics of the randomly selected subcohort are also summarised, alongside summaries

from the overall EPIC cohort from which it was sampled, to provide some indication of the representativeness of the subcohort compared with the whole of EPIC. Comparison  $p$  values were not calculated for these two groups as, because of the large sample size, even very small clinically negligible differences in the distribution of a particular characteristic are likely to be statistically significant. Prentice-weighted Cox regression models and random effects meta-analyses were used, as described in more detail in the [ESM](#), to investigate differences in the incidence of diabetes by sex and age. Crude and age-standardised incidence rates were calculated within each country.

## Results

A total of 12,403 incident cases of type 2 diabetes were ascertained and verified (Fig. 1) during 3.99 million person-years of follow-up of 340,234 EPIC participants (mean follow-up 11.7 years), excluding individuals without stored blood ( $n=109,625$ ) or without information on reported diabetes status ( $n=5,821$ ). The total number of incident cases in InterAct further excluded a total of 2,577 verified Danish cases for logistical reasons, as local sample retrieval and DNA extraction of more than the originally anticipated 2,000 cases was not feasible in the required timeframe. A random subcohort of 16,835 individuals was selected; after exclusion of 548 individuals with prevalent diabetes, 129 individuals with unknown and four with post-censoring diabetes status, 16,154 individuals were included in the subcohort for InterAct analyses. Because of the random selection, this subcohort also included a random set of 778 individuals who had developed incident diabetes during follow-up.

*Characteristics of individuals in the random subcohort* Country-specific baseline characteristics of the random subcohort were similar to those in the overall EPIC population eligible for inclusion for each country (ESM Table 2). The mean age of participants in the random subcohort was 52.5 years, a total of 38.4% were men (ESM Table 2). Average BMI and waist were 26.1 kg/m<sup>2</sup> and 86.8 cm (26.7 kg/m<sup>2</sup> and 95.4 cm in men, 25.8 kg/m<sup>2</sup> and 81.6 cm in women), respectively. A total of 57.0% of participants reported being physically inactive or moderately inactive, 34.6% were educated at secondary level or above, and 46.4% never smoked. Family history of diabetes was not ascertained in Italy, Spain, Heidelberg or Oxford; 8,307 of the 16,835 subcohort participants had information on family history; of these, the history was positive in 1,628 individuals (19.6% of those with data, or 9.7% of the subcohort).

*Characteristics of type 2 diabetes cases* Characteristics of ascertained and verified incident cases are shown in Table 2; country-specific information is provided in ESM Table 3. Including the 2,577 of the 4,632 Danish cases with available blood samples who are not part of InterAct, the overall incidence in InterAct was 3.76 per 1,000 person-years of follow-up, based on 14,980 verified diabetes cases occurring during 3,989,345 person-years. Crude incidences ranged from 1.4/1,000 person-years in French women to 7.4/1,000 (men 8.9, women 6.0) person-years in Denmark (ESM Table 3). For all analyses other than the calculation of crude and standardised incidence rates the 2,577 additional Danish cases are excluded, leading to a total of 12,403 InterAct cases (14,980–2,577=12,403). Of the 12,403 InterAct cases with an average follow-up of 6.9 years, 49.7% were men ( $n=6,165$ ). Mean baseline age and age at diagnosis were 55.6 and 62.5 years, mean BMI and waist were 29.7 kg/m<sup>2</sup> and 97.7 cm (29.4 kg/m<sup>2</sup> and 102.7 cm in men, 30.1 kg/m<sup>2</sup> and 92.8 cm in women), respectively (Table 2). A total of 30.8% of all InterAct cases reported a positive family history of diabetes (26.3% of men, 35.3% of women), excluding centres which did not obtain this information (Table 2).

*Associations of age and sex with incident type 2 diabetes* Men showed a significantly greater risk of incident diabetes than women (Fig. 2), with a pooled HR (95% CI) of 1.51 (1.39–1.64). Despite a consistent male excess of diabetes risk across all centres and countries, some heterogeneity in the effect of sex was present ( $I^2$  57%). Adjusting the centre-specific effects of sex for waist, but not BMI explained some of the heterogeneity ( $I^2$  was reduced to 33%). Diabetes incidence increased linearly with age (Fig. 3), with an overall pooled HR 1.56 (1.48; 1.64) for a 10 year age difference (1.44 [1.35–1.55]) in men and 1.64 (1.55–1.74) in women); the apparently substantial heterogeneity ( $I^2$  73% overall, 71% in men, 61% in women) occurred mainly because of the larger than average and statistically significant effect in the Bilthoven cohort (HR 2.29 in men and 2.28 in women).

Associations of measures of obesity, smoking, alcohol intake, physical activity, socioeconomic status and dietary information with diabetes are each the subject of separate InterAct reports.

## Discussion

Type 2 diabetes is an increasingly common and complex disease that clusters in families and is influenced by genetic and lifestyle factors. Despite progress in the identification of common genetic variants through genome-wide meta-analyses

**Table 2** Characteristics (mean [SD]) for continuous and % [n] for categorical variables) of 12,403 InterAct incident type 2 diabetes cases

Characteristic	Total population			Men			Women		
	n=12,403			n=6,165			n=6,238		
	Missing (%)	Mean or %	SD or n	Missing (%)	Mean or %	SD or n	Missing (%)	Mean or %	SD or n
Age (years)	0.0	55.6	7.7	0.0	55.4	7.5	0.0	55.8	7.8
Age at diagnosis (years) <sup>a</sup>	0.0	62.5	8.0	0.0	62.3	7.8	0.0	62.6	8.2
BMI (kg/m <sup>2</sup> )	0.8	29.7	4.7	0.8	29.4	4.0	0.8	30.1	5.3
Waist (cm)	7.7	97.7	12.5	8.7	102.7	10.5	6.6	92.8	12.3
Waist/hip ratio (%)	7.7	91.6	8.9	8.8	98.0	6.0	6.6	85.4	6.6
Follow-up (years)	0.0	6.9	3.3	0.0	6.9	3.3	0.0	6.8	3.3
Family history of diabetes <sup>b</sup>	1.5			1.9			1.1		
No		67.7	4,998		71.8	2,641		63.6	2,357
Yes		30.8	2,277		26.3	968		35.3	1,309
Self-reported hypertension <sup>c</sup>	3.1			4.0			2.3		
No		60.0	7,438		61.1	3,764		58.9	3,674
Yes		36.9	4,577		35.0	2,156		38.8	2,421
Self-reported hyperlipidaemia <sup>c</sup>	29.1			31.6			26.8		
No		50.2	6,231		45.2	2,784		55.3	3,447
Yes		20.6	2,557		23.3	1,435		18.0	1,122
Self-reported diabetes <sup>d</sup>	4.2			6.3			2.4		
No		33.9	2,620		35.2	1,261		32.8	1,359
Yes		61.9	4,779		58.4	2,090		64.8	2,689
Glucose-lowering drug use <sup>d</sup>	0.1			0.1			0.1		
No		47.7	3,682		47.8	1,710		47.5	1,972
Yes		52.3	4,037		52.1	1,865		52.4	2,172
Self-reported diabetes and drug use <sup>d</sup>	4.3			6.4			2.5		
No		63.3	4,887		61.6	2,204		64.7	2,683
Yes		32.4	2,505		32.0	1,144		32.8	1,361

<sup>a</sup> In 421/12,403 (3.3%) cases (235 men, 186 women), date of diagnosis was missing; for these cases date of diagnosis was imputed to be halfway between date of recruitment and end of follow-up, and this imputed date of diagnosis was used to calculate age at diagnosis

<sup>b</sup> Family history data were not collected in Italy, Spain, Heidelberg or Oxford (excluded from these summaries)

<sup>c</sup> Hypertension is 'Yes' if either 'treatment for hypertension' or 'self-reported hypertension' is positive; missing if both variables are missing (also hyperlipidaemia)

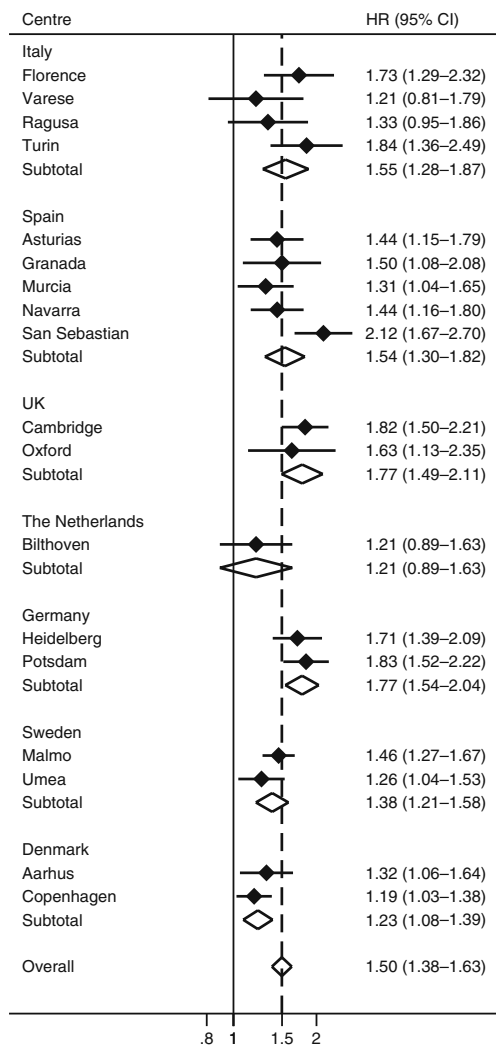
<sup>d</sup> Information on self-reported diabetes/drug use was not available in Sweden and Denmark (excluded from these summaries)

of diabetes case–control studies, current studies are lacking the power and prospective design to investigate interaction between genes and lifestyle. Effect sizes of type 2 diabetes loci identified to date are small and explain only a small proportion of familial clustering [17, 18].

Heterogeneity in effects across studies exists because of the different design and case ascertainment of cross-sectional studies contributing to earlier GWAS [19]. In addition, heterogeneity in effects of established or as yet unidentified loci between population subgroups is likely, because of the multifaceted aetiology of type 2 diabetes. Varying effects between subgroups defined by disease characteristics such as early vs late diabetes onset or with

vs without positive family history may reflect different relative contributions of genetic vs lifestyle factors. Not accounting for potential sources of heterogeneity may lead to important subgroup effects being overlooked, genetic variants not being identified and the genetic variance explained being underestimated.

Understanding differences in how genetic susceptibility translates into diabetes risk between subgroups defined by lifestyle or health behavioural factors such as obese vs lean or sedentary vs active has the potential to inform strategies for disease prevention, as previously suggested for a common variant in the *TCF7L2* gene in high-risk individuals in the Diabetes Prevention Programme [20]. However,



**Fig. 2** HRs for incident type 2 diabetes in men compared with women across InterAct centres and countries ( $I^2$  57%). France, Naples and Utrecht included women only and were excluded

existing general population cohorts with prospective data on type 2 diabetes are underpowered to systematically investigate gene–lifestyle interaction. We have therefore set up the European InterAct Study, with a nested case–cohort design that combines the advantages of a prospective cohort with the efficiency and power of a large case–control study.

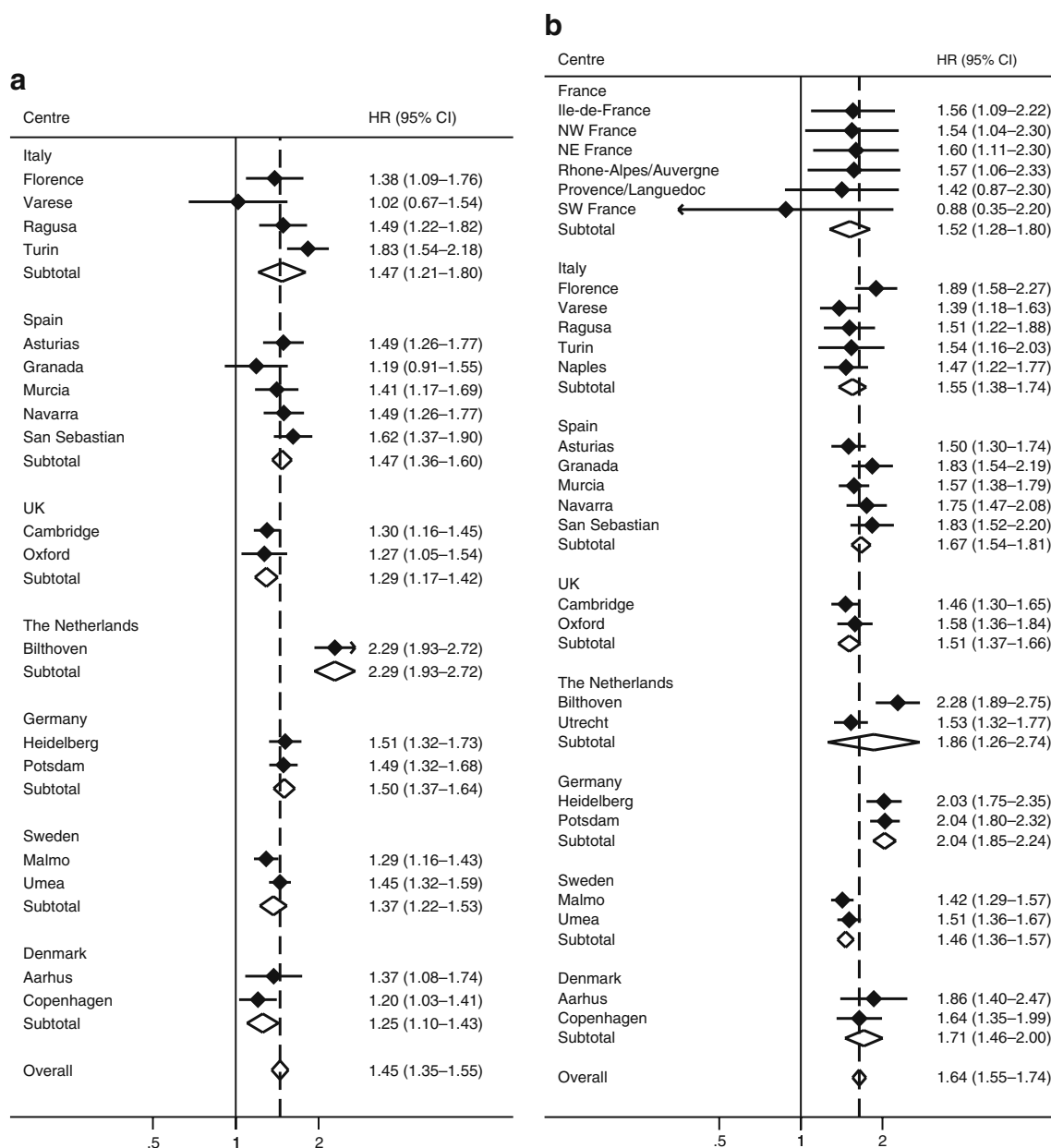
**Strength and weaknesses** The use of cases and a random cohort nested in EPIC enabled InterAct partners to jointly ascertain and verify a total of 12,403 incident type 2 diabetes cases in a relatively short time frame, exceeding the originally estimated number. Extraction of DNA in cases and in the randomly selected subcohort and storage in a comPOUND (TTP LabTech, Cambridge, UK) automatic DNA-handling system allows rapid genotyping that together with standardised baseline information on participants' clinical characteristics and health behaviour enables investi-

gation of the interaction between genes and lifestyle factors. Genome-wide data on a stratified InterAct sample of 10,000 participants provides the opportunity for discovery of as yet unidentified variants whose larger subgroup effects may have introduced sufficient heterogeneity in non-stratified conventional GWAS to prevent them from reaching the stringent significance levels required in this setting.

In addition to the prospective design that minimises systematic error introduced by recall bias, advantages of the InterAct case–cohort design include the time- and cost-efficiency and maximal sharing of resources that can be achieved by sharing of the randomly selected subcohort. In a traditional prospective cohort study such as EPIC, initially disease-free individuals are followed up over time to observe the rate of occurrence of binary clinical events (e.g. diabetes) in relation to information obtained at baseline. The major advantage is that this approach avoids the problem of recall bias, but it involves follow-up of a large study population over many years and is time consuming and expensive. This is especially true when the characterisation of possible exposures is expensive and inefficient as it is obtained in the whole cohort, only a small proportion of which go on to become cases. In a case–cohort study, efficiency is optimised by only obtaining additional exposure information for participants experiencing the outcome of interest and for members of a random sample (subcohort) selected from the entire cohort independent of the outcome. An added advantage is that this subcohort can be used as a comparison cohort for different outcomes of interest. Case–cohort studies can be designed within large cohorts with blood samples, DNA or other materials stored at baseline for later exposure measurement.

Subcohort participants were shown to be representative of EPIC participants eligible for inclusion in InterAct within each country, and thus provide an excellent opportunity for data sharing with research groups studying other disease outcomes within EPIC. Sensitivity analyses calculating HRs from unweighted Cox regression using full EPIC cohort data compared with weighted Cox regression using case–subcohort data showed that results were comparable for sex, but differed somewhat for age.

The detailed standardised evaluation of dietary and lifestyle factors on a Europe-wide scale, including assessment of nutritional biomarkers in addition to dietary self-report, optimisation of dietary data through calibration, and validation of physical activity questionnaires against objective measures, will help to address questions that have not been resolved because of inconclusive results from earlier prospective studies or intervention trials. The study of different European populations with considerable heterogeneity in dietary habits and health behaviours will increase the generalisability of any potential findings. Heterogeneity in the confounding structure across countries will help to minimise



**Fig. 3** HRs for incident type 2 diabetes per 10 years of age in (a) men, and (b) women across InterAct centres and countries ( $I^2$  71% in men, 61% in women; analysis with calendar time as the underlying timescale)

the identification of false-positive non-causal associations. The large number of cases in InterAct will be instrumental in being able to investigate dose–response effects of important exposures in more detail and inform decisions about appropriate thresholds with greater precision of estimates.

Although the possibility of examining the consistency of main genetic and interaction effects across the eight European countries will help to understand factors contributing to any potential heterogeneity, results from our InterAct participants of European descent do not allow inference about other ethnic groups with different allele frequencies and distribution and determinants of health

behaviours. We used a clinical definition of type 2 diabetes that did not rely on glucose measurement. This means that although our case definition is specific because of the verification process, InterAct rates reflect the incidence of clinical type 2 diabetes and our case number would be even larger had it been possible to identify undiagnosed diabetes cases. Likewise, we have not ruled out undiagnosed diabetes in the subcohort, which may lead to an underestimation of main effects and reduced power for interaction analyses. Despite great efforts to standardise the clinical case definition of our study, some heterogeneity exists between centres because of differences in the information



that was available locally for case ascertainment and verification. In addition, not all EPIC cohorts recruited participants from the general population [9, 10], potentially limiting the generalisability of our findings. Centre- and country-specific absolute incidence rates therefore need to be generalised and compared with caution. However, any bias through differences in the case definition between countries would need to be consistent across centres and countries to lead a false-positive association in our meta-analyses, the results of which also provide information about heterogeneity and hence, to some degree, generalisability of findings. The prevalence of important diabetes risk factors has changed since EPIC participants were recruited, with an ageing European population and greater prevalence of overweight and obesity; however, whether or not their relative importance for diabetes risk or potential interactions with genetic susceptibility has also changed remains unknown.

*Diabetes incidence by country, age and sex* Despite the increasing prevalence of type 2 diabetes and associated economic and public health burden, [21, 22] data on incidence rates from population-based European studies are scarce. Previous studies identifying clinical as well as undiagnosed asymptomatic diabetes have shown rates varying from 3.0 per 1,000 person-years in Sweden [23] to 19.1 in Southern Spain [24], with estimates for other European countries or regions lying somewhere in between [25–29]. The precision of these estimates from individual studies is naturally limited because of the relatively small cohort sizes and numbers of new cases; in addition, rates are expected to be higher compared with those based on a clinical diagnosis alone, as used in InterAct.

In contrast to many prevalence studies universally showing an increase of type 2 diabetes with age, data on differences in the incidence of type 2 diabetes in Europe are sparse. We observed a linear increase in diabetes risk by age across the eight countries included in InterAct, with an overall 56% increase in risk for a decade of age (men 45%, women 64%). This trend, in the context of an ageing European population, will lead to a rise in the incidence of diabetes if effective prevention measures are not implemented now. We also found a consistent male excess in diabetes incidence across all countries, with an overall 50% higher risk in men compared with women. Differences in waist circumference accounted for some of the heterogeneity in this association between countries. The previous literature on sex differences in incident diabetes largely supports this finding, with most [24, 26, 29–31], but not all studies [27] reporting a male excess in risk. Future InterAct work will identify factors that underlie and contribute to differences in diabetes incidence by sex and age.

The aims of the InterAct Project extend beyond the establishment of the InterAct case-cohort study and investigation of gene-lifestyle interaction using prospective observational studies. To enable InterAct partners to take gene-lifestyle interactions identified in the observational part of the study forward to a trial setting, a consortium of lifestyle intervention diabetes prevention trials ready for genotyping has been established. To better understand the impact of genetic risk information for preventive action, a Cochrane systematic review of the provision of risk information on emotion, cognition and behaviour has been completed, suggesting that the communication of DNA-based disease risk estimates may have little or no effect on behaviour, but may have a small effect on intentions to change behaviour [32].

*Conclusions* In summary, the InterAct Study is a large-scale collaborative endeavour with the potential to improve our understanding of gene-lifestyle interaction by using sufficiently powered prospective observational data including 12,403 incident type 2 diabetes cases and a random subcohort of 16,835 individuals. The establishment of a well-characterised random subcohort will aid the design and conduct of future case-cohort studies nested within EPIC cohorts.

**Acknowledgements** We thank all EPIC participants and staff for their contribution to the study. We thank staff from the Technical, Field Epidemiology and Data Functional Group Teams of the MRC Epidemiology Unit in Cambridge, UK, for carrying out sample preparation, DNA provision and quality control, genotyping and data-handling work. We specifically thank S. Dawson for coordinating the sample provision for biomarker measurements, A. Britten for coordinating DNA sample provision and genotyping of candidate markers, N. Kerrison, C. Gillson and A. Britten for data provision and genotyping quality control and M. Sims for writing the technical laboratory specification for the intermediate pathway biomarker measurements and for overseeing the laboratory work. The InterAct study received funding from the European Union (Integrated Project LSHM-CT-2006-037197 in the Framework Programme 6 of the European Community). In addition, InterAct investigators acknowledge funding from the following agencies. M. Tormo: Health Research Fund (FIS) of the Spanish Ministry of Health; the CIBER en Epidemiología y Salud Pública (CIBERESP), Spain; Murcia Regional Government (no. 6236); J. Spranger: Heisenberg-Professorship (SP716/2-1), a Clinical Research Group (KFO218/1) and Research Group (Molecular Nutrition) of the Bundesministerium für Bildung und Forschung (BMBF); Y.T. van der Schouw, J.W.J. Beulens, P.H. Peeters, I. Sluijs: verification of diabetes cases was additionally funded by NL Agency grant IGE05012 and an Incentive Grant from the Board of the UMC Utrecht; H.B. Bueno-de-Mesquita: Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (the Netherlands); M.D. Chirlaque Lopez: Health Research Fund (FIS) of the Spanish Ministry of Health; Murcia Regional Government (no. 6236); F.L. Crowe: Cancer Research UK; P. Deloukas: Wellcome Trust; L. Groop:

Swedish Research Council; G. Hallmans: The county of Västerbotten; R. Kaaks: Deutsche Krebshilfe; T.J. Key: Cancer Research UK; K.T. Khaw: MRC UK, Cancer Research UK; A. Koulman: MRC (Cambridge Lipidomics Biomarker Research Initiative); C. Navarro: Health Research Fund (FIS) of the Spanish Ministry of Health; Murcia Regional Government (no. 6236); K. Overvad: Danish Cancer Society; O. Pedersen: Faculty of Health Science, University of Aarhus, Denmark; J. R. Quirós: Asturias Regional Government; L. Rodriguez-Suarez: Asturias Regional Government; A. Tjonneland: Danish Cancer Society; R. Tumino: AIRE-ONLUS Ragusa, AVIS-Ragusa, Sicilian Regional Government; D.L. van der A, W.M.M. Verschuren: Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (the Netherlands); M. McCarthy: Wellcome Trust (083270/Z/07/Z), MRC (G0601261).

**Duality of interest** All authors declare that there is no duality of interest associated with this manuscript. I. Barroso and her spouse own stock in the companies GlaxoSmithKline and Incyte.

## Appendix

The InterAct Consortium list of authors is as follows:

C. Langenberg (MRC Epidemiology Unit, Cambridge, UK); S. Sharp (MRC Epidemiology Unit, Cambridge, UK); N.G. Forouhi (MRC Epidemiology Unit, Cambridge, UK); P.W. Franks (Lund University, Malmö, Sweden); M. B. Schulze (German Institute of Human Nutrition, Potsdam-Rehbruecke, Germany); N. Kerrison (MRC Epidemiology Unit, Cambridge, UK); U. Ekelund (MRC Epidemiology Unit, Cambridge, UK); I. Barroso (Wellcome Trust Sanger Institute, Cambridge, UK); S. Panico (Federico II University, Naples, Italy); M.J. Tormo (Department of Epidemiology, Murcia Regional Health Council, Murcia, Spain); J. Spranger (Charité University Berlin, Germany); S. Griffin (MRC Epidemiology Unit, Cambridge, UK); Y.T. van der Schouw (University Medical Center Utrecht, the Netherlands); P. Amiano (Public Health Division of Gipuzkoa, San Sebastian, Spain); E. Ardanaz (Navarre Public Health Institute, Pamplona, Spain); L. Arriola (Public Health Division of Gipuzkoa, San Sebastian, Spain); B. Balkau (INSERM, University Paris Sud, France); A. Barricarte (Epidemiology, Prevention and Promotion Health Service, Pamplona, Spain); J.W.J. Beulens (University Medical Center, Utrecht, the Netherlands); H. Boeing (German Institute of Human Nutrition, Potsdam-Rehbruecke, Germany); H.B. Bueno-de-Mesquita (National Institute for Public Health and the Environment, Bilthoven, the Netherlands); B. Buijsse (German Institute of Human Nutrition Potsdam-Rehbruecke, Germany); M.D. Chirlaque Lopez (Murcia Regional Health Authority, Spain); F. Clavel-Chapelon (INSERM, University Paris Sud, France); F.L. Crowe (University of Oxford, UK); B. de Lauzon-Guillan (INSERM, University Paris Sud, France); P. Deloukas

(Wellcome Trust Sanger Institute, Cambridge, UK); M. Dorronsoro (Public Health Division of Gipuzkoa, San Sebastian, Spain); D. Drogan (German Institute of Human Nutrition, Potsdam-Rehbruecke, Germany); P. Froguel (Imperial College London, UK); C. Gonzalez (Catalan Institute of Oncology, Barcelona, Spain); S. Grioni (Fondazione IRCCS Istituto Nazionale Tumori, Milan, Italy); L. Groop (University Hospital Scania, Malmö, Sweden); C. Groves (University of Oxford, UK); P. Hainaut (International Agency for Research of Cancer, Lyon, France); J. Halkjaer (Danish Cancer Society, Copenhagen, Denmark); G. Hallmans (Umea University, Sweden); T. Hansen (Hagedorn Research Institute, Copenhagen, Denmark); J.M. Huerta Castaño (Murcia Regional Health Authority, Spain); R. Kaaks (German Cancer Research Centre, Heidelberg, Germany); T.J. Key (University of Oxford, UK); K.T. Khaw (University of Cambridge, UK); A. Koulman (MRC Human Nutrition Research, Cambridge, UK); A. Mattiello (Federico II University, Naples, Italy); C. Navarro (Murcia Regional Health Authority, Spain); P. Nilsson (Lund University, Malmö, Sweden); T. Norat (Imperial College London, UK); K. Overvad (School of Public Health, Aarhus, Denmark); L. Palla (MRC Epidemiology Unit, Cambridge, UK); D. Palli (Cancer Research and Prevention Institute (ISPO), Florence, Italy); O. Pedersen (Hagedorn Research Institute, Copenhagen, Denmark); P.H. Peeters (University Medical Center Utrecht, the Netherlands); J.R. Quirós (Asturias Health and Health Care Council, Oviedo, Spain); A. Ramachandran (India Diabetes Research Foundation, Chennai, India); L. Rodriguez-Suarez (Asturias Health and Health Care Council, Oviedo, Spain); O. Rolandsson (Umea University, Sweden); D. Romaguera (Imperial College London, UK); I. Romieu (International Agency for Research of Cancer, Lyon, France); C. Sacerdote (Center for Cancer Prevention, Torino, Italy); M.J. Sánchez (Andalusian School of Public Health, Granada, Spain); A. Sandbaek (School of Public Health, Aarhus, Denmark); N. Slimani (International Agency for Research of Cancer, Lyon, France); I. Sluijs (University Medical Center, Utrecht, the Netherlands); A.M.W. Spijkerman (National Institute for Public Health and the Environment, Bilthoven, the Netherlands); B. Teucher (German Cancer Research Centre, Heidelberg, Germany); A. Tjonneland (Danish Cancer Society, Copenhagen, Denmark); R. Tumino (Cancer Registry and Histopathology Unit, Ragusa, Italy); D.L. van der A (National Institute for Public Health and the Environment, Bilthoven, the Netherlands); W.M.M. Verschuren (National Institute for Public Health and the Environment, Bilthoven, the Netherlands); J. Tuomilehto (University of Helsinki, Finland); E. Feskens (University of Wageningen, the Netherlands); M. McCarthy (University of Oxford, UK);

E. Riboli (Imperial College London, UK); N.J. Wareham (MRC Epidemiology Unit, Cambridge, UK).

## References

- Sladek R, Rocheleau G, Rung J et al (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 445:881–885
- Saxena R, Voight BF, Lyssenko V et al (2007) Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 316:1331–1336
- Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL et al (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316:1341–1345
- Zeggini E, Weedon MN, Lindgren CM et al (2007) Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 316:1336–1341
- Steinthorsdottir V, Thorleifsson G, Reynisdottir I et al (2007) A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat Genet* 39:770–775
- Zeggini E, Scott LJ, Saxena R et al (2008) Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* 40:638–645
- Altshuler D, Daly M (2007) Guilt beyond a reasonable doubt. *Nat Genet* 39:813–815
- Manolio TA (2009) Cohort studies and the genetics of complex disease. *Nat Genet* 41:5–6
- Bingham S, Riboli E (2004) Diet and cancer—the European Prospective Investigation into Cancer and Nutrition. *Nat Rev Cancer* 4:206–215
- Riboli E, Hunt KJ, Slimani N et al (2002) European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr* 5:1113–1124
- Slimani N, Deharveng G, CharrondiFre RU et al (1999) Structure of the standardized computerized 24-h diet recall interview used as reference method in the 22 centers participating in the EPIC project. *Comput Meth Programs Biomed* 58:251–266
- Day N, Oakes S, Luben R et al (1999) EPIC-Norfolk: study design and characteristics of the cohort. *European Prospective Investigation of Cancer. Br J Cancer* 80(Suppl 1):95–103
- Wareham NJ, Jakes RW, Rennie KL et al (2003) Validity and repeatability of a simple index derived from the short physical activity questionnaire used in the European Prospective Investigation into Cancer and Nutrition (EPIC) Study. *Public Health Nutr* 6:407–413
- Brage S, Ekelund U, Brage N et al (2007) Hierarchy of individual calibration levels for heart rate and accelerometry to measure physical activity. *J Appl Physiol* 103:682–692
- Martyn CN, Barker DJ, Osmond C (1996) Mothers' pelvic size, fetal growth, and death from stroke and coronary heart disease in men in the UK. *Lancet* 348:1264–1268
- Pischon T, Boeing H, Hoffmann K et al (2008) General and abdominal adiposity and risk of death in Europe. *N Engl J Med* 359:2105–2120
- McCarthy MI, Hirschhorn JN (2008) Genome-wide association studies: potential next steps on a genetic journey. *Hum Mol Genet* 17:R156–R165
- McCarthy MI, Zeggini E (2009) Genome-wide association studies in type 2 diabetes. *Curr Diab Rep* 9:164–171
- Timpson NJ, Lindgren CM, Weedon MN et al (2009) Adiposity-related heterogeneity in patterns of type 2 diabetes susceptibility observed in genome-wide association data. *Diabetes* 58:505–510
- Florez JC, Jablonski KA, Bayley N et al (2006) TCF7L2 polymorphisms and progression to diabetes in the Diabetes Prevention Program. *N Engl J Med* 355:241–250
- Wild S, Roglic G, Green A, Sicree R, King H (2004) Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. *Diabetes Care* 27:1047–1053
- Shaw JE, Sicree RA, Zimmet PZ (2010) Global estimates of the prevalence of diabetes for 2010 and 2030. *Diab Res Clin Pract* 87:4–14
- Jansson SP, Andersson DK, Svardsudd K (2007) Prevalence and incidence rate of diabetes mellitus in a Swedish community during 30 years of follow-up. *Diabetologia* 50:703–710
- Soriguer F, Rojo-Martinez G, Almaraz MC et al (2008) Incidence of type 2 diabetes in southern Spain (Pizarra Study). *Eur J Clin Invest* 38:126–133
- Forouhi NG, Luan J, Hennings S, Wareham NJ (2007) Incidence of type 2 diabetes in England and its association with baseline impaired fasting glucose: the Ely Study 1990–2000. *Diabet Med* 24:200–207
- Rathmann W, Strassburger K, Heier M, Holle R, Thorand B, Giani G et al (2009) Incidence of type 2 diabetes in the elderly German population and the effect of clinical and lifestyle risk factors: KORA S4/F4 cohort study. *Diabet Med* 26:1212–1219
- Bonora E, Kiechl S, Willeit J et al (2004) Population-based incidence rates and risk factors for type 2 diabetes in white individuals: the Bruneck Study. *Diabetes* 53:1782–1789
- Arteagoitia JM, Larranaga MI, Rodriguez JL, Fernandez I, Pinies JA (2003) Incidence, prevalence and coronary heart disease risk level in known type 2 diabetes: a sentinel practice network study in the Basque Country, Spain. *Diabetologia* 46:899–909
- Valdes S, Botas P, Delgado E, Alvarez F, Cadorniga FD (2007) Population-based incidence of type 2 diabetes in northern Spain: the Asturias Study. *Diabetes Care* 30:2258–2263
- Meisinger C, Thorand B, Schneider A, Stieber J, Doring A, Lowel H (2002) Sex differences in risk factors for incident type 2 diabetes mellitus: the MONICA Augsburg cohort study. *Arch Intern Med* 162:82–89
- Hippisley-Cox J, Coupland C, Robson J, Sheikh A, Brindle P (2009) Predicting risk of type 2 diabetes in England and Wales: prospective derivation and validation of QDScore. *BMJ* 338:b880
- Marteau TM, French DP, Griffin SJ et al. (2010) Effects of communicating DNA-based disease risk estimates on risk-reducing behaviours. *Cochrane Database of Systematic Reviews* 2010, Issue 10, Art. no.: CD007275. doi:10.1002/14651858.CD007275.pub2.