Title of research proposal

On the origin of healthy food consumption and physical activity: a consortium of international twin studies

Summary of research proposal

The association between modifiable risk factors for most chronic diseases and population's health are now clear. Numerous studies have reported convincing results, but at the same time a large proportion of the population still does not meet the recommendations for diet and physical activity. One of the reasons for this might be ascribed to genetic factors. Indeed, several twin studies have confirmed this, but the details of these associations are lacking. The knowledge regarding the genes effectively responsible is even scarcer. No genome-wide association studies (GWAS) have been done specifically to analyse the effect on food intake, and only one moderately large GWAS was found on physical activity behaviour.
This research aims at increasing our understanding of the relative genetic and environmental influence on healthy food consumption and physical activity by bringing together the world's data on this topic. So far, twelve twin studies from across the world, with a total of 37,177 twin pairs (74,354 individuals) have expressed their willingness to join this pooled analysis, which would be the largest ever conducted analysis on the origins of food consumption and physical activity. The heritability of these phenotypes will be calculated through a variety of variance components models and a pooled GWAS will be performed.
Discerning the origin through which healthy behaviour occurs is essential in order to develop effective personalised interventions for lifestyle change and the promotion of population's health. This is is at the core of NUTRIM's mission.

Discipline / research area

Nutrition, Nutrigenomics, Physical Activity, Behavioural Genetics, Twin studies

Details of supervisor(s):

Prof. dr. Maurice Zeegers is professor of Complex Genetics & Epidemiology. He is internationally recognised for his research on twins. He serves as head of Maastricht's School of Public Health and Primary Care and the department of Complex Genetics. He has published >200 papers and supervised >20 PhD students.
Email: m.zeegers@maastrichtuniversity.nl, Telephone: +31433882012
http://www.maastrichtuniversity.nl/web/Profile/m.zeegers.htm

Dr. Evangelia Antoniou is a psychologist and post-doctoral researcher at UM. Her current research focuses on twins' psychological development, eating behaviour and eating disorders. She is the principal investigator and coordinator of the Twins and Multiple Births Association Heritability Study (TAMBAhs).
Email: eva.antoniou@maastrichtuniversity.nl, Telephone: +31433884049
http://www.maastrichtuniversity.nl/web/Profile/eva.antoniou.htm

Dr. Annemarie Koster is associate professor at the department of social medicine. She is an epidemiologist by training and her research focuses on understanding the causes and consequences of physical activity and sedentary behaviour.
Email: a.koster@maastrichtuiversity.nl.nl, Telephone: +31433882265
http://www.maastrichtuniversity.nl/web/Profile/a.koster.htm

The student will be supervised by a group of experts with different but complementary expertise in Nutrition (AK and MZ), Physical Activity (AK), Twin Research (EA and MZ) and Statistical Genetics (EA). The student will receive multidisciplinary training in all of these topics. She will be based within the team 'Evidence-Based Nutrition' of the department of Complex Genetics at Maastricht University, the Netherlands. This productive team comprises one of the two twin research centres in the Netherlands. Here, she will work with other colleagues and PhD students that are also involved in twin research (e.g. via their projects in the East Flanders Prospective Twin Survey), other consortia (e.g. the BLEND consortium that has also harmonized dietary information) and other nutritional research projects (e.g. clinical trials and cohort studies on dietary supplementation). The student will be working in close collaboration with the department of Social Medicine, which will provide the methodological expertise regarding the assessment and interpretation of physical activity data.

Background and preliminary data

Importance of diet and physical activity on health

During the last decades, non-communicable diseases have reached epidemic proportions in all Western countries, with one main reason being the change in the population's habits concerning diet and physical activity. In fact, the so-called "obesogenic environment" that characterises modern society stimulates the intake of more calorific and palatable foods, while many technological advancements are specifically designed to provide comfort, i.e. to reduce the physical activity needed for a certain task[1].
Two meta-analyses stated the importance of introducing healthy foods in the diet in order to improve health, finding that high glycaemic load and especially high glycaemic index foods are positively associated with the

risk of type 2 diabetes, cardiovascular diseases and some forms of cancer[2], and that consumption of fruit and vegetables is inversely associated with coronary heart disease risk[3].
With regards to physical activity, it has been shown[4] that meeting the level recommended by the WHO would reduce the number of deaths worldwide due to non-communicable diseases by 5.3 million per year. Physical inactivity is therefore considered a risk factor for all major chronic diseases, comparable to smoking and obesity[4].
For all these reasons, targeting dietary intakes and physical activity behaviours is crucial in order to improve the population's health. However, despite the great deal of effort put on intervention strategies that could improve these behaviours, a large proportion of the world population does not meet dietary nor physical activity level recommendations[4].
Like many factors that "run in families", attitudes toward healthy lifestyles could be, at least in part, genetically determined. Understanding the genetic influence on modifiable risk factors is essential for the development of more effective interventions on the at-risk population, and in particular in view of a personalised approach to primary prevention.

Twins studies on dietary intake and physical activity

A twin study design, in which monozygotic (MZ) twin pairs are compared to dizygotic (DZ) twin pairs, is often used to disentangle the relative contributions of genetics and the environment on an outcome. Higher within-pair correlation in MZ than in DZ twins indicates the existence of genetic influence on the outcome. This study design has led to a wider knowledge on the heritability of dietary intake and physical activity behaviours.
Preferences for single macronutrients have been shown to be more heritable than individual foods likings[5], which are also influenced by personal experiences and the environment. Moreover, the genetic impact could follow distinct heritability patterns in men and women[6], and be age-dependent[6]. Furthermore, dietary patterns intake seems genetically determined as well. In particular, "healthy" patterns appear more genetically influenced than high-fat diets, which are essentially influenced by the "obesogenic" environment[7].
Regarding the heritability of physical activity behaviour, there is no agreement. The largest comparative twins study so far[8], which includes twins from six European countries and Australia, calculated that genetic influences range between 27% and 70%. Moreover, the genetic influence may be of different intensity between genders[8] and may decrease with age[9]. One limitation of this study is that it did not attempt to pool the results of the different studies. Also, it included twins with similar ethnicity, while not analysing the genetic influence on, for example, Asian populations. In fact, the Southeast Asia is indeed the geographic area with the least physically inactive population (17%, compared to 34.8% in

Europe)[10] and therefore it would be important to explore the relative contributions of genes and the environment in different regions of the world and across various ethnicities.

Genome-wide Association Studies (GWAS) on food intake and physical activity

There is limited knowledge on which genes are implicated in foods and macronutrients intakes, the majority of it deriving from GWAS focused on the genes-BMI association. This difference in objectives is potentially linked to a lack of power in single studies and therefore with problems in finding the genes implicated in the association. However, some of the genes related to obesity have also been associated with the intake of specific macronutrients[11].

The largest identified GWAS on the association between genes and physical activity used data from 2750 unrelated Dutch and American adults on 470,719 SNPs[12]. In this GWAS, several novel SNPs associated with physical activity are listed, together with two candidate genes (the *CYP19A1* and *LEPR* gene) that have been replicated. However, the limited study population available may have reduced its power and prevented the identification of more genes involved in the association.

It appears, then, that more large-scale studies, specifically focused on the gene-diet and gene-physical activity associations, are needed in order to clarify the relative contributions of genes and the environment on these outcomes.

Research plan

This research aims at increasing our understanding of the relative genetic and environmental influence on healthy food consumption and physical activity by bringing together the world's data on this topic.

Objective 1: To set up an infrastructure for a large twin study consortium.

Objective 2: To harmonize the study data on food frequency intake and physical activity.

Objective 3: To disentangle the genetic from the shared and unique environmental influences.

Objective 4: To integrate whole-genome analysis data to identify novel pathways.

Study recruitment

Via literature searches, the professional network of the research team and the International Network of Twin Registries, we have identified 35 twin studies with at least 100 adult twin pairs that underwent assessments on their dietary intake and/or physical activity. Twelve large twin studies (34%) from China, Sri Lanka, Finland, United Kingdom, Sweden, Belgium, Netherlands, Guinea Bissau, Australia, United States and South Korea have already agreed to participate in the project and provide their data. Table 2 in section 3d shows more information on the participating studies and principle investigators. In the unlikely event that no other studies would join the consortium, the current pooled analysis would be by far the largest ever conducted on food consumption and physical activity, with data on 37,177 twin pairs (74,354 individuals).

Data harmonization

After data cleaning, all dietary information will be classified using the hierarchical Eurocode2 food coding system[13]. On its highest level, the Eurocode2 describes common food categories such as grain, meat and beverages. The system further provides the possibility to enter information on three deeper levels of detail. Macronutrient intake (fat, carbohydrates, protein) will be calculated and all data will be equalised in grams or millilitres intake per week.

We will calculate the adherence to well-known diets such as regionally defined diets (including Mediterranean, oriental and western type) and culturally defined diets (including vegetarian, vegan and prudent diets). For this purpose, we will build indices as linear scales that incorporate selected components of the predefined diets. The total indices will be computed by summing up the level for each component in which higher values indicate higher adherence to the diet.

All physical activity data will be converted into Metabolic Equivalents (MET). We will distinguish leisure time activity from “exercise”. Leisure time includes light-intensity activities, such as walking and household jobs. Exercise means being physically active with a moderate-to-vigorous intensity for >60min/week[8,9]. Additionally, we will assess whether people meet the international exercise recommendations of 150min/week. We will analyze the duration of leisure-time activity independently of exercise. The hours of sitting/week (sedentary behaviour) will also be examined.

Data on the following confounding variables will be collected: age, gender, ethnicity, BMI, education level, and smoking.

Variance Components Analysis

Estimates of heritability will be calculated for the above-mentioned phenotypes. The Structural Equation Analysis will involve the calculation of the following components that may explain phenotypical variance: genetic, shared environmental and unique environmental. Categorical phenotypes will be assumed to follow an underlying, continuous liability, while the number of thresholds will be set depending on the categories of the phenotype. We will extend these models in order to examine possible age and sex influences. These include fitting a general-effects X-limitation model, a common-effects X-limitation model, and a X-homogeneity model. By comparing a general heterogeneity model to a common-homogeneity model we will investigate whether there are different genes that influence the phenotypes in e.g. males and females. By comparing the common-effects X-limitation model to a homogeneity model we will test whether there is any difference in the magnitude of genetic influences on phenotypes' aetiology. The resulting phenotype-specific and study-specific correlation matrices will be transformed into Fisher z values with corresponding standard errors. A meta-analysis will be used to pool these z-scores across studies. Heterogeneity between studies will be modelled by using a random intercept, and potential sources of heterogeneity (such as the prevalence of environmental exposures) will be investigated via meta-regression techniques. The pooled z-values will be converted back to correlations and then to variance components for ease of interpretation. All analyses will be performed using the statistical software OpenMx (version 3.7) and R (version 3.2.4). Power analyses[14] show that with 37,177 twin pairs even variance components that contribute <5% to the total phenotypical variance can be identified. The power of the study will increase when more twin studies join the consortium.

Integration of GWAS data

For those studies that have GWAS data available, we will impute genotypes based on sequencing data to enrich the dataset to its maximum capacity. We will then calculate the heritability for each genetic variant. These percentages will be ranked and displayed in Manhattan plots. We will list the genotypes with the top10, top25 and top100 heritabilities. Potential overlap between variant locations and known bioinformatic features will be investigated by using ICSC test browser, HaploReg, RegulomeDB and the Gene Expression Atlas.

Time line, feasibility

As this project is data driven, there is a low risk of not delivering the proposed tasks in time. Theoretically, the data transfer or data harmonization could take longer than planned. This is however unlikely, given that the PIs of the twin studies have collaborated before on different phenotypes. In addition, the supervisory team has harmonized dietary data using the Eurocode2 in consortium studies before. The thesis will produce at least 5 academic papers.

Table 1: Gantt chart of project milestones

Milestones	Year 1		Year 2		Year 3		Year 4	
1: Harmonize study data	■							
2: Data-linkage of study data	■	■						
3: Variance Components Analyses (VCA) on dietary food items and food groups (Paper 1)		■	■					
4. VCA on macronutrient intake, total calorie intake and calorie expenditure (Paper 2)			■	■				
5. VCA on the adherence to dietary patterns (Paper 3)					■			
4: Statistical analyses on physical activity data (Paper 4)					■	■		
6: Integration of GWAS data (Paper 5)						■	■	■
7: Completion of PhD thesis								■

*Year 1 includes twin studies recruitment, data transfer and agreements, writing of protocol and statistical analyses plan, and training in twin modelling and meta-analyses.

Originality and / or innovative elements of the topic and approach

This proposal aims to fill critical gaps in current scientific knowledge and to understand the origin of healthy food consumption and physical activity. Our approach offers several innovative methods to advance the field including:

- Detailed exposure assessment of twelve twin studies from different populations from across the world
- Very large sample size, using at least 37,177 twin pairs (74,354 individuals)
- The most advanced statistical techniques to discern nature versus nurture
- No pooled analysis has been done before on these associations
- The integration of GWAS data that will investigate many gene-diet and gene-physical activity associations for the first time and on a massive scale

Relevance

The economic and societal costs of healthy eating and physical activity are substantial. Apart from its large impact on population health (outlined in paragraph 2a), it is estimated that, by 2030, the health-care costs for chronic, obesity-related diseases will increase by $48-66 billion yearly in the USA. To this, the cost of loss of productivity should be added, estimated to be $390-580 billion in 2008[15].
Discerning the origins through which healthy behaviour occurs is essential in order to develop effective personalised interventions for lifestyle change. If more phenotypical variance is explained by genetic influence, then lifestyle interventions may benefit from genetic stratification. If shared-environment is more important, family-tailored interventions may be more effective.

Literature references

1. Lake A. and Townshend T. (2006) Obesogenic environments: exploring the built and food environments. *The Journal of the Royal Society for the Promotion of Health*. **126**(6),262-267.
2. Barclay A.W., Petocz P., McMillan-Price J., Flood V.M., Prvan T., Mitchell P. and Brand-Miller J.C. (2008) Glycemic index, glycemic load, and chronic disease risk—a metaanalysis of observational studies. *American Journal of Clinical Nutrition*. **87**, 627–37
3. Dauchet L., Amouyel P., Hercberg S. and Dallongeville J. (2006) Fruit and Vegetable Consumption and Risk of Coronary Heart Disease: A Meta-Analysis of Cohort Studies. *Journal of Nutrition*. **136**, 2588–2593
4. Lee I-M., Shiroma E.J., Lobelo F., Puska P., Blair S.N. and Katzmarzyk P.T. for the Lancet Physical Activity Series Working Group (2012) Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *Lancet*. **380**, 219–29. http://dx.doi.org/10.1016/S0140-6736(12)61031-9.
5. Reed D.R., Bachmanov A.A., Beauchamp G.K., Tordoff M.G. and Arlen Price R. (1997) Heritable Variation in Food Preferences and Their Contribution to Obesity. *Behavioural Genetics*. **27**(4), 373–387.
6. Hasselbalch A.L., Heitmann B.L., Kyvik K.O. and Sørensen T.I. (2008) Studies of twins indicate that genetics influence dietary intake. *Journal of Nutrition*. **138**(12), 2406-12. doi: 10.3945/jn.108.087668.
7. Pallister T., Spector T.D. and Menni C. (2014) Twin studies advance the understanding of gene–environment interplay in human nutrigenomics. *Nutrition Research Reviews*. 1-10. doi:10.1017/S095442241400016X
8. Stubbe J.H., Boomsma D.I., Vink J.M., Cornes B.K., Martin N.G., *et al*. (2006) Genetic Influences on Exercise Participation in 37.051 Twin Pairs from Seven Countries. *PLoS ONE*. **1**(1): e22. doi:10.1371/journal.pone.0000022
9. Vink J.M., Boomsma D.I., Medland S.E., de Moor M.H.M., Stubbe J.H., Cornes B.K., Martin N.G., Skytthea A., Kyvik K.O., Rose R.J., Kujala U.M., Kaprio J., Harris J.R., Pedersen N.L., Cherkas L., Spector T.D. and de Geus E.J.C. (2011) Variance Components Models for Physical Activity With Age as Modifier: A Comparative Twin Study in Seven Countries. *Twin Research and Human Genetics*. **14**(1) 25–34. doi:10.1375/twin.14.1.25
10. Hallal P.C., Andersen L.B., Bull F.C., Guthold R., Haskell W. and Ekelund U. for the Lancet Physical Activity Series Working Group (2012) Global physical activity levels: surveillance progress, pitfalls,

and prospects. *Lancet*. **380**, 247-57.

11. Bauer F., Elbers C.C., Adan R.A.H., Loos R.J.F., Onland-Moret N.C., Grobbee D.E., van Vliet-Ostaptchouk J.V., Wijmenga C. and van der Schouw Y.T. (2009) Obesity genes identified in genome-wide association studies are associated with adiposity measures and potentially with nutrient-specific food preference. *American Journal of Clinical Nutrition*. **90**, 951–9.
12. De Moor M.H., Liu Y.J., BoomsmaD.I., Li J., Hamilton J.J., Hottenga J.J., Levy S., Liu X.G., Pei Y.F., Posthuma D., Recker R.R., Sullivan P.F., Wang L., Willemseng G., Yan H., De Geus E.J., Deng H.W. (2009) Genome-wide association study of exercise behavior in Dutch and American adults. *Medicine & Science in Sports & Exercise*. **41**(10), 1887-1895.
13. Poortvliet, E.J., Klensin, J.C. and Kohlmeier, L. (1991) *Rationale Document for the Eurocode Food Coding System*. Institute for Social Medicine and Epidemiology, Berlin.
14. Visscher P.M. (2004) Power of The Classical Twin Design Revisited. *Twin Research*. **7**(5), 505-512.
15. Wang Y.C., McPherson K., Marsh T., Gortmaker S.L. and Brown M. (2011) Health and economic burden of the projected obesity trends in the USA and the UK. *Lancet*. **378**, 815-825.

16.

Table 2: Overview of twin registries that so far agreed to collaborate

Study name	Country	Collaborator	#MZ pairs	#DZ pairs	Dietary data	Physical Activity data	Genetic data
The Swedish Twin Study of Adults: Genes and Environments-STAGE	Sweden	Dr. Patrik Magnusson	2321	1479	X	X	X
TwinsUK	United Kingdom	Dr. Victoria Vazquez	1426	822	X	X	X
The South Korean Twin Registry	South Korea	Dr. Joohon Sung	416 (in 2006)*	183 (in 2006)*	X	X	
Chinese National Twin Registry	China	Dr. Wenjing Gao, Prof. Liming LI	250*	250*	X	X	
Finnish Twin Registry-FinnTwin 16	Finland	Prof. Jaakko Kaprio	1408	2980	X	X	
Sri Lankan Twin Registry	Sri Lanka	Dr. Kaushalya Jayaweera	1691	2274	X	X	
University of Washington Twin Registry	United States of America	Prof. Glen Duncan	3832	3387	-	X	

Bandim Health Project Twin Registry	Guinea Bissau	Dr. Morten Bjerregaard-Andersen	*	*	-	X	
East Flanders Prospective Twin Survey	Belgium	Dr. Catherine Derom	1000*	2000*	-	X	X
Netherlands Twin Registry*	The Netherlands	Prof. dr. Eco de Geus	3636*	7272*	-	X	X
Queensland Twin Registry	Australia	Prof. Nicholas Martin	*	*	-	X	
Twins & Sisters Study	Australia	Dr. Janine Rochelle Lam	330*	220*	X	X	X

*Actual number may differ, as the response rate to the assessment of dietary intake and physical activity was still unknown at the time of submitting this proposal.

** The NTR will provide correlation matrices instead of individual data