MISSION-T2D

Multiscale Immune System SImulator for the Onset of Type 2 Diabetes integrating genetic, metabolic and nutritional data

Work Package 7

Deliverable 7.3

Report on analysis and validation of the integrated model results

Document Information

Grant Agreement	Nº	600803	Acronym	MISSION-T2D
Full Title	Multiscale Immune System SImulator for the Onset of Type 2 Diabetes integrating genetic, metabolic and nutritional data			
Project URL	http://www.mission-t2d.eu			
EU Project Officer	Name	Dr. Adina Ratoi		

Deliverable	No	7.3	Title	Report on analysis and validation of the integrated model results
Work package	No	7	Title	Clinical guidance, results analysis, T2D risk assessment and validation

Date of delivery	Contractual	01 May 2016	Actual	03 May 2016
Status	Version 1.1		Final	03.05.2016

Nature	Prototype		Report	X	Dissemination		Other	

Dissemination level	Consortium+EU	X
	Public	

Target Group (If Public)		Society (in general)	
Specialized research communities		Health care enterprises	
Health care professionals		Citizens and Public Authorities	

Responsible Author	Name	Albert de Graaf	Partner	TNO
	Email	albert.degraaf@tno.nl		

Version Log			
Issue Date	Version	Author (Name)	Partner
28.04.2016	1.0	Albert de Graaf	TNO
02.05.2016	1.1	Filippo Castiglione	CNR
03.05.2016	1.2	Albert de Graaf	TNO

Executive Summary	This document reports on Task 7.4: Assist the modeling WPs in preprocessing the experimental data to “model-ready” parameters. A status update on progress of the TNO P4 Health studies and on data availability from the studies identified in Tasks 7.1 - 7.3 is given. Data use requests for additional datasets for validation of the MISSION-T2D models were filed and access to several additional datasets was granted. Data were pre-processed and partially stored in the phenotype database dbNP for easy retrieval by partners. The procedure for dbNP access and data download is described. Data preparation of the Whitehall II cohort data transferred to MISSION-T2D partners was performed and is described.
Keywords	data provision, model validation datasets, intervention study datasets, type 2 (pre)diabetes subgroups, data preparation, data storage, data retrieval

Contents

1. Deliverable Description

Deliverable 7.3 is the outcome of Task 7.4 named: “Assist the modeling WPs in preprocessing the experimental data to “model-ready” parameters”.
The specific objective of Task 7.4 was to store all data of task 7.1-7.3 in a dedicated database (the phenotype database, www.dbnp.org), which allows web-based access for all relevant partners. In addition, preprocessing and formatting to fit modeling was to be performed. Also, efforts were continues to secure more data for MISSION-T2D. This deliverable reports on the status of the TNO health care innovation program, on the current state of availability of data for use in the MISSION-T2D project, as well as on data storage and preprocessing of the most relevant data from WP7.

2. Deliverable Results

2.1. General remarks

In Tasks 7.1 and 7.2, datasets were identified that contain data from challenge tests and intervention studies. Task 7.3 should expand on these tasks by identifying datasets that are more specifically focused on actual healthcare treatment of T2D patients.
We re-iterate that at the inception of MISSION-T2D it was anticipated that such data would primarily come from studies executed in the framework of a TNO health care innovation program (P4 medicine field trial) that was being set up at the time of submission of the MISSION-T2D proposal. With considerable delay, several TNO P4 health field trials had begun during the previous reporting period, and they have considerably progressed since then. In the P4 Hillegom field lab, 35 of the aimed-for total number of 60 diabetes patients are enrolled. Unfortunately, data have not yet become available. Another study, the QuaLiFY Field Lab, involving personalized dietary advice to type-2 diabetes patients based on several measured health parameters along with markers for nutrition status and genetic SNPs, was finished in October 2015 and a report was prepared. Results showed no effects of the intervention. The TNO-internal employee pilot study, P4@TNO, involving self-measurement and dry blood spot analysis had finished and was analized. Results indicated that self-monitoring of health parameters had a positive effect on body weight. After significant delays were incurred due to the complicated ethics procedures (Medical Ethics Committee approval had/has to be obtained in every country from which participants originate), the Nutrition Researcher Cohort Field Lab, featuring the use of do-it-yourself devices, filling out online-questionnaires and sample collection with supplied kits for the analysis of various health parameters, commenced in January 2015 and is still on-going. To date there are approximately 200 participants from countries where ethical approval has been obtained. Some preliminary data have become available.

Summaries of these 4 studies are described in section 2.2.

Since the P4 Health studies so far provided very few data on type-2 diabetes patients, in Task 7.4 during the present reporting period we continued the efforts to obtain data use approval for several studies already identified in Tasks 7.1 and 7.2. Notably we were successful with an important new dataset, i.e., the DIOGENES study. Some additional datasets were also shared

with the MISSION-T2D Consortium. Continuing the work in Task 7.3, data from the most relevant studies was extracted for pre-assessment and stored in the database dbNP (databasing infrastructure developed by TNO for nutritional systems biology studies (http://www.dbnp.org) that is ideal to store data in a querable format; introduced in Deliverable 7.1) for easy retrieval by MISSION-T2D partners. The procedure of using dbNP to extract pre-processed data formatted to fit modelling is summarized in Annex 2.

A status update is given in section 2.3 while section 2.4 provides some background on the Diogenes study.

During the first project Review Meeting in May 2014, the Whitehall II cohort in particular was suggested for validation of the MISSION-T2D models. After Whitehall II data use permission was obtained shortly before the end of the previous reporting period, in the present period the data was pre-processed for validation by the MISSION-T2D models. A short report detailing these challenging activities is given in section 2.5, while a description of the data is given in Annex 1. The approach taken for the validation using this and other data, in which TNO actively participated, is described in Deliverable 4.4 and also in CNR's Deliverable 6.3 "Report on the validation of the computational model and refinement of the integrated model" (due PM36).

With the described progress, which all taken together resulted in the information supplied in Deliverables 7.1 to 7.3, we expect to have adequately addressed the reviewers' recommendation to "demonstrate the availability of appropriate data for model validation" that resulted from the second project Review Meeting hold in Brussels in June 2015.

2.2. *Status of the TNO P4 health care innovation program*

As detailed in the previous reporting period, the TNO P4 health care innovation program comprises 4 different studies. Two studies are concerned with diabetes patients (the P4 Hillegom study and the QuaLiFY field lab), one study with TNO employees (the P4@TNO study), and one study with nutrition researchers (the Nutrition Research Cohort (NRC) study).

The current state of progress of these studies is addressed below.

2.2.1 Progress of the TNO P4 health studies

2.2.1.1 P4 Hillegom Study

Title: P9607 P4 approach in newly diagnosed type 2 diabetics in the Hillegom Field Lab

Study setup: Rationale, Objective, Study design, Study population, Personalized diagnosis, Intervention amd personalized treatment and Main study parameters/endpoints were all described in Deliverable 7.2.

Progress: The study has started Q4 of 2014. So far, thirty-five of the required total of 60 patients have entered the study and underwent the 3 month intervention. Follow-up data are not yet available.

2.2.1.2 P4@TNO Pilot Study

Title: P9608 P4@TNO pilot study: potential of do-it-yourself devices for obtaining personal health data

Study setup: Rationale, Objective, Study design, Study population, Intervention, Self-monitoring devices (Medisana) and data recording and Main study parameters/endpoints were all described in Deliverable 7.2.

Progress: The study was completed in Q4 2014. Data were analyzed and a report has been prepared. Results were presented as oral presentation: de Hoogh IM, Boessen R, Boorsma A, Bobeldijk I, Pasman WJ; "TNO acting as a field lab for health measurements; effects on health and behavior"; NASO (The Netherlands Association for the Study of Obesity) Scientific Spring Meeting, April 13, 2016, Utrecht (NL) with Dissemination funding from MISSION-T2D.

Results: 33 TNO employees started the study and 31 subjects completed the study. From the n-of-1 statistic approach a significant effect on weight loss (slope = -0.1 kg) and BMI (slope = -0.27) was found. Energy expenditure increased significantly with 41.2 kcal/day (sd=22.7) during the study. No effect was found on blood glucose or blood pressure. The results of the user experience questionnaire showed that 86% of the participants felt that study participation improved the insight in their health status, and 51% of the participants agreed that the use of do-it-yourself devices helped them to improve their health.

Conclusion: Self-monitoring of health parameters had a positive effect on body weight, which could be explained by increased energy expenditure. For most participants self-monitoring increased awareness of their health status. Self-monitoring has the potential to increase awareness of own health status and thereby serve as a motivator for behaviour change.

2.2.1.3 QuaLiFY Field Lab

Title: P9618 – Personalized Dietary Advice in diabetes type 2

Study setup: Rationale, Objective, Study design, Study population, Intervention and Main study parameters/endpoints were all described in Deliverable 7.2.

Progress: Twenty-eight subjects were enrolled into the study. The study was completed in Q4 2015. Data were analyzed and a report has been prepared.

Access to scientifically valid data and knowledge rules relevant to personalized nutritional products and services is provided via the QuaLiFY Server Platform (QSP) or Quisper (http://www.qualify-fp7.eu/qualify-server-platform).

Efficacy results:

No effect of the Personalized Dietary Advice was found on the established markers of diabetes type 2, namely fasting glucose, fasting c-peptide and HbA1c.

No effect was found of the Personalized Dietary Advice intervention on blood pressure (systolic and diastolic) in subjects.

No effect of the Personalized Dietary Advice intervention was found on the cholesterol levels of subjects. There is also no effect of the Personalized Dietary Advice intervention found on the HDL, LDL or triglyceride levels of subjects.
No effect of the Personalized Dietary Advice intervention was found on anthropometric measures (weight, BMI, waist circumference, hip circumference, fat percentage) of subjects.
No effect of the Personalized Dietary Advice intervention was found on biomarkers of food intake (cholesterol, plasma carotenoids, plasma vitamins, plasma fatty acids, omega 3 index) of subjects.
No effect of the Personalized Dietary Advice intervention was found on scores for vitality, energy, motivation or resilience of subjects, as measured by the Vita-16.
No effect of the Personalized Dietary Advice intervention was found on scores on any of the scales of the RAND-36 (physical functioning, role limitation due to physical health, role limitations due to emotional problems, vitality, emotional well-being, social functioning, pain, general health) of subjects.
No effect of the Personalized Dietary Advice intervention was found on scores on physical activity or sitting behavior of subjects, as measured by the IPAQ.

Qualitative evaluation results: According to dieticians, both the intervention and control group had a positive effect of the study, because both were provided with feedback based on their measurements, for example on fasting glucose and cholesterol measurements.
The 'MijnEetmeter.nl' system (both available via a mobile app and a website) was found to be much more intuitive as compared to OPEN by both the dieticians and the participants.
According to the dieticians, the activity tracker motivated participants to be more physically active.
In general, the study portal was considered easy and intuitive to work with by dieticians. Especially, the feedback in the form of graphs of the various measurements was appreciated.
The dieticians experienced some problems with the cholesterol meter at the start of the study. However, after a few measurements this became easier. The hip circumference measurement was more difficult with obese participants and therefore considered unreliable. The dieticians experienced no problems with carrying out the glucose, blood pressure and waist circumference measurement.
The sample collection kits provided by Vitas (finger prick blood) and GeneticLab (buccal-swabs for DNA) were considered to be easy-to-use. The finger prick blood kit provided by SwissAnalysis was found to be less user-friendly by both dieticians and participants and should be improved for use in future studies.
The biggest problem according to the dieticians about the study was the lack of information about the different measurements that were used in the feedback to the subjects, like healthy cut-off values and how these can be influenced by a healthy diet.
Dieticians found the advice texts that were given to the participants to specific and sometimes not in line with their professional opinion.
Most participants rated their study participation as pleasant and agreed that the study participations increased the insight in their own health and resulted in more healthy behaviour.

Conclusions:

This study did not show any advantageous or adverse effects of Personalized Dietary Advice on any of the established markers of nutritional and health status in diabetes type 2 patients, as compared to regular care.
The feedback to the participants based on the (blood) measurements was considered very motivating.

2.2.1.4 Nutritional Researcher Cohort Field Lab

Title: P9616 NRC n250 study

Study setup: Rationale, Objective, Study design, Study population, Intervention and Main study parameters/endpoints were all described in Deliverable 7.2.

Progress:
Due to the new nature of the NRC procedures (complete "do-it-yourself" analysis, personal data upload, anonymized research, inclusion of personal genetic testing), intense discussions on ethics were necessary.
Despite the initial delay caused by complicated ethics procedures in several European countries (Medical ethics Committee approval had/has to be obtained in every country from which participants originate), the NRC partners exchanged lessons learned in dealing with the ethics aspects and succeeded in achieving ethics approval of the protocol by the local ethical committees in Finland, Czech Republic, United Kingdom, Ireland, the Netherlands, Austria, Italy, Switzerland, Spain, Belgium and France. The protocol was also submitted in Germany but is still pending approval.

The following tools have been assessed, with some illustrative results detailed below:
- The Vitas dried blood spot testing;
- The SwissAnalysis clinical chemistry package;
- A dried-blood spot based metabolome analysis;
- The TNO "do-it-yourself" OGTT (oral glucose tolerance test);
- The TNO "do it yourself" microbiome assay;
- The FatSecret food intake quantification;
- The 23andMe 1M SNP analysis connected to the SafeCape advice system, in combination with a series of specifically designed bioinformatics applications.

Actual assessment of the tools commenced in January 2015 and is still on-going. To date approximately 200 participants are included from countries where ethical approval has been obtained (see figure 2.2.1).

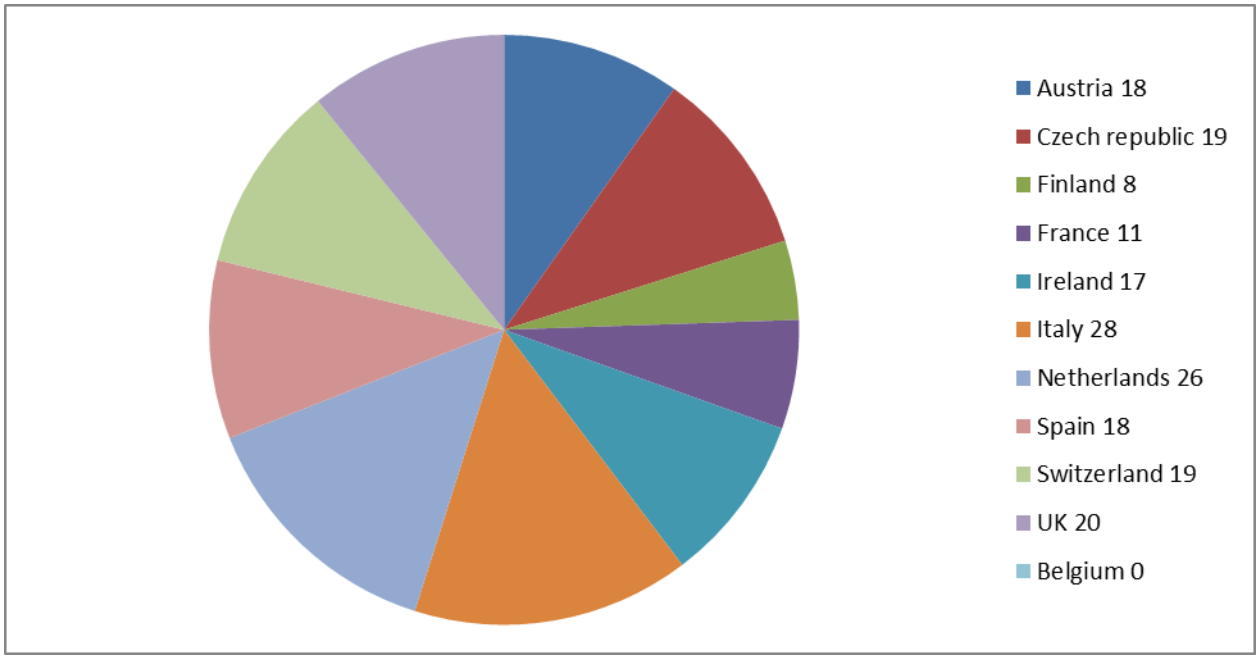


Figure 2.2.1. Number of participants recruited for the NRC 250 study where a number of self-assessment tools were evaluated.

Illustrative intermediate results of different self-assessment tools:

The Vitas dried blood spot testing, was used for total fatty acids and amino acid analysis. An example of obtained results is shown in Figure 2.2.2.

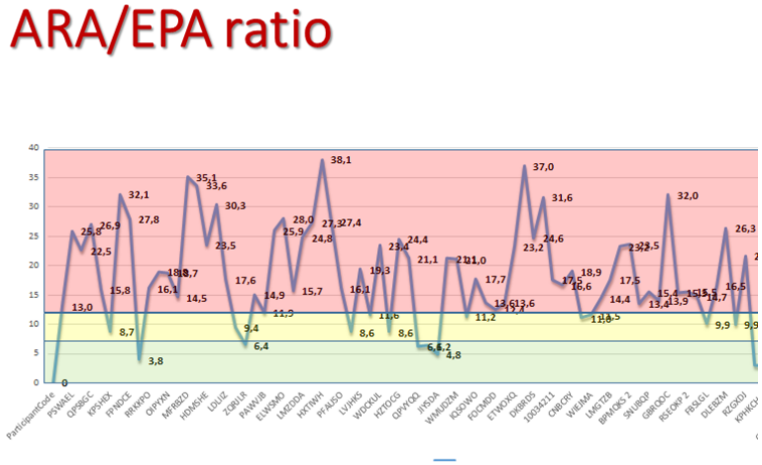


Figure 2.2.2. Illustration of measured omega fatty acids ratio of Arachidonic acid (ARA) and Eicosapentaenoic acid (EPA). The ratios are an indicator of cardio vascular disease risk, in this figure high risk indicated by red, medium risk indicated by yellow and low risk indicated by green zone. The NRC participants indicated in this figure show a wide distribution in the calculated ARA/EPA ratio. These ratios are used for personalized food advice with respect to omega 3 and omega 6 fatty acid intake recommendations.

The SwissAnalysis clinical chemistry package appeared to require much more blood then was achieveable for most participants in a Do-it-yourself setting. Swiss Analysis is currently developing a more user-friendly sampling device and procedures in a different project, that may be used for self-monitoring in the future.

The do-it-yourself (DiY) OGTT set up by TNO was used for sampling of dry blood spots (DBS) (see Figure 2.2.3.) for insulin and c-peptide analysis.

Figure 2.2.3. DiY OGTT based on sample collection of paper cards.

The quality of the blood samples was good; 95% of the samples were suitable for analysis. TNO further validated the insulin assay in a separate project. The results showed that insulin analysis from DBS is not reliable and therefore insulin was not determined. For c-peptide, in the validation study, a very good correlation was found between c-peptide determined from dry blood spots and from venous blood collection. It was therefore decided that in the future only c-peptide will be determined from the blood spots.

For the TNO microbiome assay, a simple DiY fecal sampling procedure was tested and combined with sequence analysis of fecal samples. Although sample collection and analysis were successful (Figure 2.2.4), at present, the results cannot be used for personalized food advice. There are as yet no reference values in the literature with respect to the occurrence and abundance of the different phyla. More data has to be collected and more studies on the relationship between diet and microbiota composition are necessary.

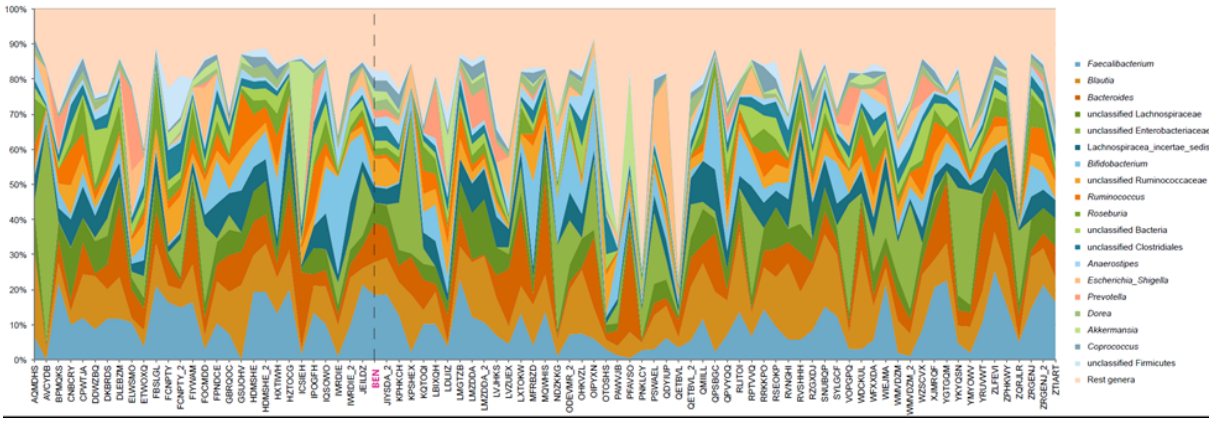


Figure 2.2.4. Most abundant gut microbial phyla as detected in the NRC250 faecal samples.

For the dietary intake registration, the Food Intake Quantification procedures developed by the FP7 Project Food4Me could not be implemented in time due to internal discussions within Food4Me on external access rights for commercial and non-commercial purposes. The open access mobile app "FatSecret" was used as an alternative. Although food intake in the NRC is asked for only 3 days (2 week days and 1 weekend day) every 3 months, compliance in the NRC is particular low. Up to now 100 percent of the participants complied to enter their food intake at the start of the NRC 250 study, before the sample collection, but only approximately 50 % of the participants enter their data regularly as described in the protocol. Whereas compliance is a known problem for food intake diaries there is also the concern of under- or over-estimation of the amount of food consumed. Another problem is that keeping a food diary directly will influence food intake behavior of the participants. After considering various alternatives, it was concluded that the most advanced option of food intake quantification would be to make use of nutritional status markers that can be measured via blood, saliva or urine.

Conclusions

Sharing of lessons learned and discussions between all countries involved in submission of new 'participatory' type of study protocols resulted in efficient knowledge sharing and as a final result, approval of NRC study protocols in 12 countries within 12 months.

The NRC cohort, spread over 12 European countries, counts more than 200 participants with nutrition research background and served as a beta testing site for testing several do-it-yourself sampling methods and devices for health status monitoring.

Sampling and measurement methods developed by project partners Vitas, Swiss Analysis and TNO were assessed by the NRC participants. Feedback from the participants was used to improve the user friendliness and other characteristics of the methods as well as visualisation of the results to the user. Several of the services were integrated into the Quisper platform (Vitas, Swiss Analysis).

The data obtained from the measurements were used as input for a personalised food advice tool. Automated Personal Dietary Advice was developed based on Food4Me algorithms.

2.3. Update on datasets identified in Tasks 7.1 and 7.2 for use in MISSION-T2D.

This section updates on the data availability of the datasets identified in Task 7.1 and 7.2. For ease of comparison, the same tables as in D7.1 and D7.2 are shown, with colour markings indicating status changes as described with the tables.

2.3.1. Stress response datasets

The studies with TNO participation are listed in Table 2.3.1.
Bioclaims (Southampton): Data use request was answered positively.
Bioclaims (Krakow): Data use request was answered positively.
PhenFlex: Data has become available. Data were already in dbNP.
PhenFlex2 and Nutritech are additional studies offering challenge test data that have finished since the previous reporting period. Data of these studies are already in dbNP. Data use has been requested.
DIOGENES: after continued attempts, our data-sharing request for the DIOGENES study was finally granted and additional data (exceeding the data that was already in dbNP in the previous reporting period) was transferred to TNO. This dataset is in the process of being decoded and pre-processed for analysis. DIOGENES was a large study, referred to in many publications. Therefore, some background information on this important study is given in section 2.4.

Table 2.3.1 – Update on studies with TNO participation to provide stress response curves of (pre)diabetes subgroups. Yellow fields indicate changes since the previous reporting period.

OGTT, Oral Glucose Tolerance Test; OLTT, Oral Lipid Tolerance Test; CVD, Cardiovascular disease; dbNP, the nutritional phenotype database.

Study name	Subjects	Type of stress response	Dynamic Metabolic variables 1: ClinChem (glucose, insulin, TG, FFA, glycerol, TC, HDL-C) 2: metabolomics	Dynamic Inflammatory variables 0: none 1: CRP, (IL-6), (TNFa) 2: endothelial function (sICAM, sVCAM, P-selectin) 3: cytokine panel	Data availability status 1: requested 2: confirmed N.A.: not available	Data location 1: dbNP 2: within TNO 3: external
EU-BIOCLAIMS Southampton	50 obese, 50 lean	OLTT	1	3	**2**	**1**
EU-BIOCLAIMS Krakow	N=150, 50 healthy vs 100 obese	OLTT, OGTT	1	0	**2**	3

PhenFlex	20 healthy, 20 diabetic	OLTT, OGTT	1,2	1,2,3	**2**	1
DIOGENES	773 (pre) diabetic subtypes	OGTT	1	1	**2**	1
CordioPrev	1000 CVD patients, diabetic subtypes	OGTT	1	1	**N.A.**	3
Phenflex2	100 subjects, 50% M/F, 10 (pre-) diabetic groups	MMTT	1,2	0	1	1
NutriTech	72 subjects (38 women, 34 men, 59 years old BMI 29)	OGTT MMTT MMTT+PA	1,2	1,2,3	1	1

2.3.2. Challenge test datasets from studies without TNO involvement

Table 2.3.2 lists the challenge test datasets from consortia without TNO involvement. Attempts to obtain data use permission for NUGENOB, LIPGENE and MECHE continued to remain unsuccessful due to complete unresponsiveness to our repeated data availability requests. The data of Lipgene and MECHE however were uploaded in dbNP in the framework of cooperations outside of MISSION-T2D. The data from WHITEHALL II were transferred to TNO and preprocessed for analysis. Details are given in section 2.5.

Table 2.3.2– Update on data availability from large studies without TNO involvement to provide stress response curves of (pre)diabetes subgroups. Yellow fields indicate changes since the previous reporting period.

OGTT, Oral Glucose Tolerance Test; OLTT, Oral Lipid Tolerance Test; MetS, Metabolic Syndrome; dbNP, the nutritional phenotype database.

Study name	Subjects	Type of stress response	Dynamic Metabolic variables 1: ClinChem (glucose, insulin, TG, FFA, glycerol, TC, HDL-C) 2: metabolomics	Dynamic Inflammatory variables 0: none 1: CRP, (IL-6), (TNFa) 2: endothelial function (sICAM, sVCAM, P-selectin) 3: cytokine panel	Data availability status 0: identified 1: requested 2: confirmed	Data location 1: dbNP 2: within TNO 3: external
NUGENOB	711 obese, 119 lean	OLTT	1	0	1	3
LIPGENE	subcohort 74 MetS	OLTT	0	2	1	1
MECHE	200 healthy	OLTT, OGTT	2	1	1	1
WHITEHALL II	10,308	OGTT	1	0	**2**	2
Oded Shaham	22 healthy, 25 prediab.	OGTT	2	0	0	1

2.3.3. TNO in-house challenge test datasets

In Task 7.1 and 7.2, twenty potentially useful TNO in-house human intervention studies carried out over the past 15 years were identified, data use requests were issued where necessary, and studies were ranked according to their relevance for MISSION-T2D. In the past period, most of the studies with highest relevance for MISSION-T2D were uploaded in dbNP. Table 2.3.3 shows the updated information for the challenge test data.

Table 2.3.3 – Update on data availability from TNO in-house human studies to provide stress response curves of (pre)diabetes subgroups. Yellow fields indicate changes since the previous reporting period.

MTT, Meal Tolerance Test; OGTT, Oral Glucose Tolerance Test; OLTT, Oral Lipid Tolerance Test; MetS, Metabolic Syndrome; IGT, impaired glucose tolerance; dbNP, the nutritional phenotype database.

TNO Study #	Subjects	Type of stress response	Dynamic Metabolic variables 1: ClinChem (glucose, insulin, TG, FFA, glycerol, TC, HDL-C) 2: metabolomics	Dynamic Inflammatory variables 0: none 1: CRP, (IL-6), (TNFa) 2: endothelial function (sICAM, sVCAM, P-selectin) 3: cytokine panel	Data availability status 1: available N.A.: not availabl	Data location 1: dbNP 2: within TNO 3: external	Data pre-assessed 1: first priority 2: lower priority 0: unsuited
462	26 healthy	MTT	1	0	**1**	2	**0**
473	10 healthy	vaccine	0	1	**N.A.**	2	
3884	24 MetS	MTT	1	1, 2	**1**	2	**0**
4616	12 IGT	OGTT	1	0	**N.A.**	2	
5328	10 healthy, 10 obese	MTT, OGTT	1	1,2	**1**	2	**2**
5805	10 healthy	MTT, OGTT	1	0	**1**	2	**2**
6281	9 lean, 9 obese	MTT, OGTT	1	0	**1**	2	**2**
6957	36 MetS	OGTT, OLTT	1	1	**1**	2	**1 see 2.4**
7261	12 obese	MTT	1	0	**1**	1	**1 see 2.4**
7573	24 healthy	MTT	1	0	**1**	2	**2**
8374	16 lean, 16 obese	exercise	1	3	**1**	1	**1 see 2.4**
8600	22 lean	LPS inflammation	1	1,2	**1**	1	**1 see 2.4**

8738	8 lean, 8 obese	OGTT+ OLTT	1	3	**1**	1	**1 see 2.4**
8749	24 healthy	mental stress	0	3	**1**	2	**0**
9218	10 MetS, 10 lean	OGTT, OLTT	1	0	**1**	1	**1 see 2.4**
9334	24 healthy	alcohol	1	0	**1**	2	**0**
9466	10 healthy	OLTT	2	3	**N.A.**	2	

2.3.4. Intervention study datasets from consortia in which TNO participates.

Datasets identified are the same as in 2.3.1 and identical comments apply. Updates are shown in Table 2.3.4.

Table 2.3.4 – Update on data availability from intervention studies with TNO participation to provide data of (pre)diabetes sub-phenotypes. Yellow fields indicate changes since the previous reporting period.

CVD, Cardiovascular disease; FA, fatty acid; dbNP, the nutritional phenotype database.

Study name	Subjects	Type of intervention	Metabolic variables 1: ClinChem (glucose, insulin, TG, FFA, glycerol, TC, HDL-C) 2: metabolomics	Inflammatory variables 0: none 1: CRP, (IL-6), (TNFa) 2: endothelial function (sICAM, sVCAM, P-selectin) 3: cytokine panel	Data availability status 1: requested 2: confirmed N.A.: not available	Data location 1: dbNP 2: within TNO 3: external
EU-BIOCLAIMS Southampton	50 obese, 50 lean	12 wk n-3 FA	1	3	2	**1**
EU-BIOCLAIMS Krakow	N=150, 50 healthy vs 100 obese	RCT 3 mo isocaloric diet +/- Epax	1	2	2	3
DIOGENES	773 (pre) diabetic subtypes	weight loss, then 4 diets	1	1	2	2

CordioPrev	1000 CVD patients, diabetic subtypes	mediterranean diet	1	1	**N.A.**	3
NutriTech	72 subjects (38 women, 34 men, 59 years old BMI 29)	Caloric restriction	1,2	1,2,3	1	1

2.3.5. Datasets from intervention studies without TNO involvement

Datasets identified are the same as in 2.3.2 and are shown in Table 2.3.5, now indicating the interventions (no interventions in MECHE and Oded Shaham). The same comments as in 2.3.2 apply. NUGENOB and LIPGENE were unresponsive; WHITEHALL II data were transferred to TNO for analysis in MISSION-T2D and pre-processed for analysis (see 2.5).

Table 2.3.5 – Update on data availability from large intervention studies without TNO involvement to provide data of (pre)diabetes sub-phenotypes. Yellow fields indicate changes since the previous reporting period.

MetS, Metabolic Syndrome; dbNP, the nutritional phenotype database.

Study name	Subjects	Type of Intervention	Metabolic variables 1: ClinChem (glucose, insulin, TG, FFA, glycerol, TC, HDL-C) 2: metabolomics	Inflammatory variables 0: none 1: CRP, (IL-6), (TNFa) 2: endothelial function (sICAM, sVCAM, P-selectin) 3: cytokine panel	Data availability status 0: identified 1: requested 2: confirmed	Data location 1: dbNP 2: within TNO 3: external
NUGENOB	711 obese, 119 lean	10 week, hypocaloric low- vs. high-fat diet	1	1	1	3
LIPGENE	417 MetS	12 week, 4 different dietary fat compositions, isocaloric	1	1	1	**1**

WHITEHALL II	10,308	lifestyle over long time	1	1	**2**	**2**

2.3.6. TNO in-house intervention study datasets

Datasets identified are the same as in 2.3.3 and are shown in Table 2.3.6, now indicating the interventions (some TNO studies only have challenge tests and no interventions, or the other way round). The same comments as in 2.3.3 apply.

Table 2.3.6 – Update on data availability from TNO in-house human intervention studies to provide data of (pre)diabetes subphenotypes. Yellow fields indicate changes since the previous reporting period.

MetS, Metabolic Syndrome; IGT, impaired glucose tolerance; MCFA, medium-chain fatty acid; LCFA, long-chain fatty acid; HFD, high-fat diet; dbNP, the nutritional phenotype database.

TNO Study #	Subjects	Type of intervention	Metabolic variables 1: ClinChem (glucose, insulin, TG, FFA, glycerol, TC, HDL-C) 2: metabolomics	Inflammatory variables 0: none 1: CRP, (IL-6), (TNFa) 2: endothelial function (sICAM, sVCAM, P-selectin) 3: cytokine panel	Data availability status 1: available N.A.: not available	Data location 1: dbNP 2: within TNO 3: external	Data pre-assessed 1: first priority 2: lower priority 0: unsuited
473	10 healthy	10 wk, 2 different FA supplement	0	1	**1**	2	**0**
3884	24 MetS	4 wk, 40g alcohol/day	1	1, 2	**1**	2	**0**
4616	12 IGT	4 wk antidiabetic peptide	1	0	**N.A.**	2	
5067	40 healthy	4 wk antidiabetic peptides	1	1	**N.A.**	2	
5328	10 healthy,	3 wk, 40g alcohol/day	1	1,2	**1**	2	**2**

	10 obese						
5805	10 healthy	4 wk, 32g/day alcohol	1	0	**1**	2	**2**
6281	9 lean, 9 obese	3 wk, 35g/day alcohol	1	0	**1**	2	**2**
6957	36 MetS	5 wk, 3 dairy products, 1 TNO-foodmix	1	3	**N.A.** **1**	2	 **1** **see 2.4**
7261	12 obese	3 wk MCFA vs LCFA supplement	1	1	**1**	**1**	**1** **see 2.4**
7348	50 pre-diabetic	weight maintenance	2	0	**1**	2	**0**
7573	24 healthy	3 wk, 26 g/day alcohol	1	0	**1**	2	**2**
8189	8 obese	10 wk, energy restriction diet	1 ?	1 ? gut microbiota	**N.A.**	2	
8318	66 low-grade inflamm	4 wk, 5 different probiotics	0	1	**N.A.**	2	
8374	16 lean, 16 obese	4 wk, 2 doses vegetables, then 4 wk, energy-restricted diet	1	3	**1**	**1**	**1** **see 2.4**
8749	24 healthy	2 wk 26g/day alcohol	1	3	**1**	2	**0**
9218	10 MetS, 10 lean	4 wk weight maintenance (MetS) or HFD	1	gut microbiota composition	**1**	**1**	**1** **see 2.4**
9466	10 healthy	3 wk, high vs no dairy	2	3	**N.A.**	2	

2.4. Background information on the DIOGENES study.

Data from the DIOGENES study were made available to MISSION-T2D and were transferred to TNO. While some data (mostly clinical chemistry and metabolomics) was already available in dbNP, the additional data now acquired allow a much more systemic perspective since also variables related to mental health are included. Below some background information on DIOGENES is given. A complete description of the study design and methods can be found in Larsen et al., (2010) The Diet, Obesity and Genes (Diogenes) Dietary Study in eight European countries – a comprehensive design for long-term intervention, Obesity Reviews 11, 76-91.

2.4.1 DIOGENES brief study description.

Rationale: The importance of the composition of a diet for the prevention and management of obesity is debated. Ad libitum consumption of low-fat diets results in short-term weight loss, and low-carbohydrate, high-protein, and high-fat diets (e.g., the Atkins diet) may result in substantial weight loss as compared with that achieved with other types of diets. However, the weight loss is generally not sustained beyond 1 year. Greater weight loss with low-carbohydrate diets may be ascribed to the satiating effects of high protein content, and there is increasing interest in the efficacy of diets that have a high protein content with a moderate carbohydrate and fat content.
A diet with a low glycemic index may have beneficial effects on body weight and body composition and on certain risk factors in overweight persons, but the effectiveness of ad libitum consumption of low-glycemic-index diets for weight control is controversial.

Objective: The Diet, Obesity, and Genes (Diogenes) study is a pan-European, multicenter, randomized, dietary-intervention study designed to assess the efficacy of moderate-fat diets that vary in protein content and glycemic index for preventing weight regain and obesity-related risk factors after weight loss.

Study design: Randomized controlled intervention, two-by-two factorial design.

Study population: The study was conducted across eight European countries and involved 891 families with at least one overweight or obese parent and one healthy child (irrespective of weight).

Intervention: Adult participants underwent an initial eight-week low-calorie diet to lose weight, with participants losing an average of 11.0kg. Families with minimum one parent attaining a weight loss of 8% or more, then were randomized to follow one of five different dietary regimes for six months:
Group 1: Low protein/low GI
Group 2: Low protein/high GI
Group 3: High protein/low GI
Group 4: High protein/high GI
Group 5: Control (national dietary guidelines) diet

All the diets were low in fat (less than 30% of energy consumed). Each group had people drop out, but the higher-protein, lower-GI diet group had the lowest dropout rate and also regained the least weight over the six months of the study.

Main study parameters/endpoints: The primary outcome measures for adults were body-weight-loss maintenance (kg), change in body composition as assessed by Dual energy X-ray Absorptiometry (DXA) or Bioelectrical Impedance Analysis (BIA), number of subjects maintaining more than 5% or more than 10% of their initial weight loss, and dropout rate during the dietary intervention. The secondary outcome measures were changes in abdominal fat mass (measured by DXA, waist circumference and sagittal diameter), risk factors for type 2 diabetes and cardiovascular disease, changes in appetite and satiety hormones, physical activity and identification and quantification of fat tissue messenger ribonucleicacid (mRNA), and certain metabolite, peptide and protein biomarkers from blood. In addition, assessments of biological (genetic profiles, measurements of basal metabolic rate (BMR) and free-living energy expenditure, assessment of physical activity, etc.) and psychological features (questionnaires covering issues such as appetite and food preferences, health promoting behaviour, attitudes towards eating, social support, etc.) that determined the families' and the individuals' responses to the dietary intervention were performed.

Results: A total of 1209 adults were screened (mean age, 41 years; body-mass index 34), of whom 938 entered the low-calorie-diet phase of the study. A total of 773 participants who completed that phase were randomly assigned to one of the five maintenance diets; 548 completed the intervention (71%). Fewer participants in the high-protein and the low-glycemic-index groups than in the low-protein-high-glycemic-index group dropped out of the study (26.4% and 25.6%, respectively, vs. 37.4%; P=0.02 and P=0.01 for the respective comparisons). The mean initial weight loss with the low-calorie diet was 11.0 kg. In the analysis of participants who completed the study, only the low-protein-high-glycemic-index diet was associated with subsequent significant weight regain (1.67 kg; 95% confidence interval [CI], 0.48 to 2.87). In an intention-to-treat analysis, the weight regain was 0.93 kg less (95% CI, 0.31 to 1.55) in the groups assigned to a high-protein diet than in those assigned to a low-protein diet (P=0.003) and 0.95 kg less (95% CI, 0.33 to 1.57) in the groups assigned to a low-glycemic-index diet than in those assigned to a high-glycemic-index diet (P=0.003). The analysis involving participants who completed the intervention produced similar results. The groups did not differ significantly with respect to diet-related adverse events.

Publication(s):
The above main results were published in Larsen et al. (2010) Diets with high or low protein content and glycemic index for weight-loss maintenance. N Eng J Med 363 (22) 2102-2113. Many more aspects of the DIOGENES study were, and continue to be, published in a large set of publications, including (but not limited to) the ones listed below.

- Stroeve J.H.M., Saccenti E., Bouwman J., Dane A., Strassburg K., Vervoort J., Hankemeier T., Astrup A., Smilde A.K., Van Ommen B., Saris W.H.M. (2016) Weight loss predictability by plasma metabolic signatures in adults with obesity and morbid obesity of the DiOGenes study, Obesity 24: 379–388.

- Navas-Carretero S., Holst C., Saris W.H., van Baak M.A., Jebb S.A., Kafatos A., Papadaki A., Pfeiffer A.F.H., Handjieva-Darlenska T., Hlavaty P., Stender S., Larsen T.M., Astrup A., Martinez J.A. (2016) The Impact of Gender and Protein Intake on the Success of Weight Maintenance and Associated Cardiovascular Risk Benefits, Independent of the Mode of Food Provision: The DiOGenes Randomized Trial, Journal of the American College of Nutrition 35: 20-30
- Larsen S.C., Ängquist L., Østergaard J.N., Ahluwalia T.S., Vimaleswaran K.S., Roswall N., Mortensen L.M., Nielsen B.M., Tjønneland A., Wareham N.J., Palli D., Masala G., Saris W.H., van der A D.L., Boer J.M.A., Feskens E.J.M., Boeing H., Jakobsen M.U., Loos R.J.F., Sørensen T.I.A., Overvad K. (2016) Intake of total and subgroups of fat minimally affect the associations between selected single nucleotide polymorphisms in the PPARγ pathway and changes in anthropometry among European adults from cohorts of the DiOGenes study, Journal of Nutrition 146: 603-611
- Engberink M.F., Geleijnse J.M., Bakker S.J.L., Larsen T.M., Handjieva-Darlesnka T., Kafatos A., Martinez J.A., Pfeiffer A.F.H., Kunešová M., Jebb S.A., Holst C., Astrup A., Saris W.H.M., Brink E.J., Van Baak M.A. (2015) Effect of a high-protein diet on maintenance of blood pressure levels achieved after initial weight loss: The DiOGenes randomized study, Journal of Human Hypertension 29: 58-63
- Aller E.E.J.G., Larsen T.M., Claus H., Lindroos A.K., Kafatos A., Pfeiffer A., Martinez J.A., Handjieva-Darlenska T., Kunesova M., Stender S., Saris W.H.M., Astrup A., Van Baak M.A. (2014) Weight loss maintenance in overweight subjects on ad libitum diets with high or low protein content and glycemic index: The DIOGENES trial 12-month results, International Journal of Obesity 38 : 1511-1517
- Brahe L.K., Ängquist L., Larsen L.H., Vimaleswaran K.S., Hager J., Viguerie N., Loos R.J.F., Handjieva-Darlenska T., Jebb S.A., Hlavaty P., Larsen T.M., Martinez J.A., Papadaki A., Pfeiffer A.F.H., Van Baak M.A., Sorensen T.I.A., Holst C., Langin D., Astrup A., Saris W.H.M. (2013), Influence of SNPs in nutrient-sensitive candidate genes and gene-diet interactions on blood lipids: The DiOGenes study, British Journal of Nutrition 110: 790-796
- McConnon A., Horgan G.W., Lawton C., Stubbs J., Shepherd R., Astrup A., Handjieva-Darlenska T., Kunešová M., Larsen T.M., Lindroos A.K., Martinez J.A., Papadaki A., Pfeiffer A.F.H., Van Baak M.A., Raats M.M. (2013) Experience and acceptability of diets of varying protein content and glycemic index in an obese cohort: Results from the Diogenes trial, European Journal of Clinical Nutrition 67 : 990 -995
- Stocks T., Ängquist L., Hager J., Charon C., Holst C., Martinez J.A., Saris W.H.M., Astrup A., Sørensen T.I.A., Larsen L.H. (2013) TFAP2B-dietary protein and glycemic index interactions and weight maintenance after weight loss in the DiOGenes trial, Human Heredity 75: 213-219
- Kunešová M., Hlavatý P., Tvrzická E., Staňková B., Kalousková P., Viguerie N., Larsen T.M., Van Baak M.A., Jebb S.A., Martinez J.A., Pfeiffer A.F.H., Kafatos A., Handjieva-Darlenska T., Hill M., Langin D., Žák A., Astrup A., Saris W.H.M. (2012) Fatty acid composition of adipose tissue triafter weight loss and weight maintenance: The DIOGENES study, Physiological Research 61 : 597-607
- Vimaleswaran K.S., Ängquist L., Hansen R.D., Van Der A. D.L., Bouatia-Naji N., Holst C., Tjønneland A., Overvad K., Jakobsen M.U., Boeing H., Meidtner K., Palli D., Masala G., Saris W.H.M., Feskens E.J.M., Wareham N.J., Sørensen T.I.A., Loos R.J.F. (2012)

Association between FTO variant and change in body weight and its interaction with dietary factors: The diogenes study, Obesity 20 : 1669-1674

- Larsen L.H., Ängquist L., Vimaleswaran K.S., Hager J., Viguerie N., Loos R.J.F., Handjieva-Darlenska T., Jebb S.A., Kunešova M., Larsen T.M., Martinez J.A., Papadaki A., Pfeiffer A.F.H., Van Baak M.A., Sørensen T.I.A., Holst C., Langin D., Astrup A., Saris W.H.M. (2012) Analyses of single nucleotide polymorphisms in selected nutrientsensitive genes in weight-regain prevention: The DIOGENES study, American Journal of Clinical Nutrition 95: 1254-1260
- McConnon A., Raats M., Astrup A., Bajzová M., Handjieva-Darlenska T., Lindroos A.K., Martinez J.A., Larson T.M., Papadaki A., Pfeiffer A., van Baak M.A., Shepherd R. (2012) Application of the Theory of Planned Behaviour to weight control in an overweight cohort. Results from a pan-European dietary intervention trial (DiOGenes), Appetite 58: 313-318
- Gögebakan O., Kohl A., Osterhoff M.A., Van Baak M.A., Jebb S.A., Papadaki A., Martinez J.A., Handjieva-Darlenska T., Hlavaty P., Weickert M.O., Holst C., Saris W.H.M., Astrup A., Pfeiffer A.F.H. (2011) Effects of weight loss and long-term weight maintenance with diets varying in protein and glycemic index on cardiovascular risk factors: The diet, obesity, and genes (diogenes) study: A randomized, controlled trial, Circulation 124: 2829-2838
- Goyenechea E., Holst C., van Baak M.A., Saris W.H.M., Jebb S., Kafatos A., Pfeiffer A., Handjiev S., Hlavaty P., Stender S., Larsen T.M., Astrup A., Martinez J.A. (2011) Effects of different protein content and glycaemic index of ad libitum diets on diabetes risk factors in overweight adults: The DIOGenes multicentre, randomized, dietary intervention trial, Diabetes/Metabolism Research and Reviews 27 : 705 - 716
- Halkjær J., Olsen A., Overvad K., Jakobsen M.U., Boeing H., Buijsse B., Palli D., Tognon G., Du H., Van Der A D.L., Forouhi N.G., Wareham N.J., Feskens E.J.M., Sørensen T.I.A., Tjønneland A. (2011) Intake of total, animal and plant protein and subsequent changes in weight or waist circumference in European men and women: The Diogenes project, International Journal of Obesity 35: 1104-1113
- Du H., Vimaleswaran K.S., Ängquist L., Hansen R.D., van der A D.L., Holst C., Tjønneland A., Overvad K., Jakobsen M.U., Boeing H., Meidtner K., Palli D., Masala G., Bouatia-Naji N., Saris W.H.M., Feskens E.J.M., Wareham N.J., Sørensen T.I.A., Loos R.J.F. (2011) Genetic polymorphisms in the hypothalamic pathway in relation to subsequent weight change - the DiOGenes study, PLoS ONE 6 , e17436
- Papadaki A., Linardakis M., Larsen T.M., Van Baak M.A., Lindroos A.K., Pfeiffer A.F.H., Martinez J.A., Handjieva-Darlenska T., Kunesová M., Holst C., Astrup A., Saris W.H.M., Kafatos A. (2010) The effect of protein and glycemic index on children's body composition: The DiOGenes randomized study, Pediatrics 126: e1143-e1152
- Handjieva-Darlenska T., Handjiev S., Larsen T.M., Van Baak M.A., Jebb S., Papadaki A., Pfeiffer A.F.H., Martinez J.A., Kunesova M., Holst C., Saris W.H.M., Astrup A. (2010) Initial weight loss on an 800-kcal diet as a predictor of weight loss success after 8 weeks: The Diogenes study, European Journal of Clinical Nutrition 64: 994-999
- Moore C.S., Lindroos A.K., Kreutzer M., Larsen T.M., Astrup A., Van Baak M.A., Handjieva-Darlenska T., Hlavaty P., Kafatos A., Kohl A., Martinez J.A., Monsheimer S., Jebb S.A. (2010) Dietary strategy to manipulate ad libitum macronutrient intake, and glycaemic index, across eight European countries in the Diogenes Study, Obesity Reviews 11: 67-75

- Larsen T.M., Van Baak M., Pfeiffer A., Martinez J.A., Handjieva-Darlenska T., Kunešová M., Pihlsgård M., Stender, S., Holst C., Saris W.H.M., Astrup A. (2010) High or Low Protein Content and Glycemic Index for Weight-Loss Maintenance, New England Journal of Medicine 363, 2102-213
- Larsen T.M., Dalskov S., Van Baak M., Jebb S., Kafatos A., Pfeiffer A., Martinez J.A., Handjieva-Darlenska T., Kunešová M., Holst C., Saris W.H.M., Astrup A. (2010) The diet, obesity and genes (diogenes) dietary study in eight European countries - A comprehensive design for long-term intervention, Obesity Reviews 11 : 76-91
- Buijsse B., Feskens E.J.M., Schulze M.B., Forouhi N.G., Wareham N.J., Sharp S., Palli D., Tognon G., Halkjaer J., Tjønneland A., Jakobsen M.U., Overvad K., Van Der A D.L., Du H., Sørensen T.I.A., Boeing H. (2009) Fruit and vegetable intakes and subsequent changes in body weight in European populations: Results from the project on Diet, Obesity, and Genes (DiOGenes), American Journal of Clinical Nutrition 90: 202-209
- Ueland Ø., Saris W. (2006) Obesity and food technology: The DiOGenes european integrated project , Food Science and Technology 20: 22-24 :
- Saris W.H.M. (2005) DiOGenes: An integrated multidisciplinary approach to the obesity problem in Europe, Nutrition Bulletin 30: 188-193
- Saris W.H.M., Harper A. (2005) Short report: DiOGenes - A multidisciplinary offensive focused on the obesity epidemic, Obesity Reviews 6: 175-176
- Green J.H., Mela D., Hill R.N. (2005) EU integrated Project on diet, obesity and genes (DiOGenes) and opportunities for food product innovation, Agro Food Industry Hi-Tech 16 : 23-24

2.5. *Preprocessing of the Whitehall II dataset for model validation.*

2.5.1 Whitehall II.

Whitehall II is a landmark longitudinal, prospective cohort study of British civil servants, with the explicit intention of examining reasons for the social gradient in health and disease in men and extending the research to include women. The phases are summarised below (from Marmot and Brunner, 2005). MISSION-T2D received data from Phase 3 and Phase 5.

Phase	Dates	Type	Participants	Age (y)
1	1985-1988	Screening/questionnaire	10308	35 to 55
2	1989-1990	Questionnaire	8133	37 to 60
3	1991-1993	Screening/questionnaire	8637	39 to 64
4	1995-1996	Questionnaire	8629	42 to 65
5	1997-1999	Screening/questionnaire	7830	45 to 69
6	2001	Questionnaire	7344	48 to 71
7	2003-2004	Screening/questionnaire	6914	50 to 74
8	2006	Questionnaire	7173	53 to 76
9	2008-2009	Screening/questionnaire	?	55 to 80

Below is an overview of the data collected (from Marmot and Brunner, 2005).

Demographic data

Socioeconomic data

- Education
- Household composition
- Income
- Financial assets
- Work + work change (retirement)

Area-level indicators

- Deprivation
- Classification of area

Psychosocial/work exposure

- Effort–reward
- Demand–control
- Social support
- Social networks

Health behaviours

- Smoking
- Alcohol
- Diet—food frequency
- Physical activity

CVD

- WHO chest pain
- Details of cardiovascular disease (CVD) symptoms, investigations, and treatment

General health (subjective)

- Self-rated health
- Well-being
- Longstanding illness
- Hospital admissions
- Medications
- Musculoskeletal conditions
- Quality of life (SF-36)

Mental health (subjective)

- General Health Questionnaire (GHQ) (anxiety, depression)
- Center for Epidemiologic Studies Depression Scale (CESD)
- SF-36, Activities of daily living (ADL), Instrumental ADL

Health outcomes (objective)

- Sickness absence
- Myocardial infarction and coronary surgery
- Stroke
- Clinical depression
- CVD/CHD mortality
- Other cause-specific mortality
- Mortality

The phases of medical examination (from Marmot and Brunner, 2005):

	Phase 1	Phase 3	Phase 5	Phase 7
Examination	Weight, height, BP	Weight, height, waist–hip ratio, BP	Weight, height, waist–hip ratio, BP	Weight, height, waist–hip ratio, BP, walking speed, spirometry
Neuroendocrine	Blood pressure reactivity[b]		Heart rate variability[b] hypothalamic–pituitary–adrenal axis measurements[b]	Heart rate variability Salivary cortisol diurnal rhythm
Subclinical cardiovascular disease	ECG: Minnesota codes ECG left ventricular mass (LVM)	ECG: Minnesota codes ECG LVM	ECG: Minnesota codes ECG LVM Ultrasound (US) measures of endothelial dysfunction[b] Carotid artery wall thickness/distensibility[b] MRI (white matter lesions)[b]	ECG: Minnesota codes ECG LVM US measures of endothelial dysfunction[b] Carotid artery wall thickness/distensibility
Lipids	Total cholesterol, apoA1, and apoB[b]	Total + HDL cholesterol apoA1 and B, Lp(a) Triglycerides Cholesterol ester fatty acids[b]	Total + HDL cholesterol Triglycerides	Total + HDL cholesterol Triglycerides
Carbohydrate metabolism		Fasting and post-load glucose and insulin	Fasting and post-load glucose and insulin	HbA1c, fasting, and post-load glucose and insulin
Genotype		DNA isolation	APOE	CRP, TNF, UCP2, MAO further genotyping
Haemostatic and other	Fibrinogen[b] Factor VIIc[b]	Fibrinogen, IL-6, CRP Factor VIIc von Willebrand factor PAI − 1[b] plasma β-carotene	Fibrinogen, IL-6[b], CRP[b], SAA[b] Viscosity D-dimer[b]	Fibrinogen, IL-6, CRP

Data received:

- Whitehall II data set with data from phase 3 (S3) and phase 5 (S5). A few variables for S1, S7 and S9 were received.
- Data dictionary
- Meaning of values (data formats)

The data were not complete. At a later stage we requested for additional data:

- Age at DM diagnosis (old WHO)
- First visit where DM was diagnosed (old WHO)
- Last visit before DM was diagnosed (old WHO)
- Follow up time for inc dm from S1 (yrs, old WHO)
- Follow up time for inc dm from S3 (yrs, old WHO)
- Year diabetes was diagnosed (S9)

In the Annex a list with variables received together with some decriptives of the data are included.

Problems encountered during data pre-processing to fit modelling:

- Meaning of values (data formats) were not complete. Especially the formats for longstanding illnesses were not delivered. Also a translation of codes about medicine taken was not delivered. At our request the additional information was sent.
- Variable 'Year to diabetes': Some subjects have a value for "age at DM diagnosis (dmagedia)" and some subjects have a number for "Year diabetes was diagnosed (S9)"(jdiabyr). There are subjects present, which have values for both variables. In these cases, we cross-checked by calculating the age at diabetes diagnosis by subtracting birth year (yob_c + 1900) from jdiabyr. In quite a large number of cases, the two ages do not match. The largest discrepancy observed is 16.3. We are wondering what could have caused these inconsistencies (perhaps different diabetes criteria?) and we do not know which number to trust more. This information is crucial for to map time trajectories until diabetes. We were unable to extract information on how to deal exactly with the inconsistencies in the "time to diabetes" calculation from a number of publications we checked. Upon addressing the Whitehall II team with this problem, the answer we received did not allow us to resolve the issue. We therefore were unable to map the data on a timeline "years to diabetes".
- Variable 'Food energy': We encountered some weird values in the food energy data (derived from Food Frequency Questionnaires by the Whitehall II team). For instance for Phase 5, food energy kcal, the minimum value seen was 40.4 and the maximum value 7079.8. Both seem very unrealistic values.
- Potential wrong conversion of data. We encountered some cases where totally unrealistic values for variables were found which later turned out to be an issue of decimal points/commas or thousands separator point/comma.
- For some variables, different questionnaires were used at S3 vs. S5 making direct comparison difficult and leaving room for systematic errors when comparing the two phases. This was especially relevant for physical activity, where we used the

categorization info at S3 (only 3 categories) to convert the S3 values to MET-hrs/week units reported at S5 (derived from a more detailed questionnaire) to hopefully solve the issue. Also, the categorization of medicine at S5 was completely different from the one at S3. Since we only used a flag for anti-diabetic medication (including insulin) we expect no serious problems at the present stage.

Final data set for first analysis and model validation:
The following data selection criteria were applied:

- No diabetes medication used
- Subjects needed to be fasted

A total of 6226 subjects fulfilled these criteria. A reduced dataset was extracted that included the variables from S3 & S5 as indicated in Table 4.1 shown in the Annex. These variables relate to the (partially latent) variables in the high-level model (see Deliverable 4.4):

- Food intake pattern
- Physical activity pattern
- BMI
- Age
- Fasting glucose
- Beta cell function
- Insulin sensitivity
- Fasting insulin
- Inflammation (adiponectin, CRP, IL6,
- Chronic stress
- Tissue damage
- Gut health
- Food quality
- Sleep/Meditation/Relaxation

References

1. Marmot and Brunner, International Journal of Epidemiology 2005, 34: 251–256, doi:10.1093/ije/dyh372.

3. Deliverable Conclusions

Task 7.4.

This deliverable concludes the work done in WP7 “Clinical guidance, results analysis, T2D risk assessment and validation” of MISSION-T2D. Making up the overall balance, despite delays involved with the establishment of the TNO P4 health studies and the limited response obtained on external study data use requests, a considerable number of highly relevant datasets for MISSION-T2D have been secured for the purpose of WP7. These include most importantly, data for 2 phases of the Whitehall II landmark cohort, data from the important DIOGENES study, six TNO datasets, as well as several other relevant datasets. Extensive data pre-processing and formatting to suit modeling purposes has been performed. Together these studies offer a truly rich source of data from healthy and type 2 diabetic patients on metabolism and inflammation, resulting from a wide range of interventions and challenge tests, at time scales ranging from minutes to several years.

4. Annexes

4.1 Annex 1: Variables and some descriptives of the Whitehall II data

Table 4.1 below shows variables present in the received data at study phases S3 and S5.

Table 4.1. Inventory of Whitehall II dataset contents transferred to TNO. Only variables for phase S3 (red highlighting and S5 (blue highlighting) were made available for MISSION-T2D. Variables with column identifier highlighted in orange were used for initial model validation.

Whitehall Variable at S3	Column in data sheet		Whitehall Variable at S5	Column in data sheet	Loads on model variable
Date of participation (S3)	A		Date of participation (S5)	C	
Date of received questionnaire (S3)	B		Date of questionnaire completion (S5)	D	
Date of screening (S3)	E		Date of screening (S5)	F	
Anonymised Whitehall ID	G				
Sex	H				
Ethnicity (1=white, 2=non-white)	I				

Year of birth	J				
Participation status (S3)	K		Participation (S5)	BV	
Age at participation date (xphdate) (S3d)	L		Age at participation date (tphdate) (S5d)	BW	
Age at quest completion (S3d)	M		Age at quest completion (S5d)	BX	Age
Incident DM at end of S3 (old WHO,cumulat from S1)	N		Incident DM at end of S5 (old WHO, cumulat from S1)	BY	
Known DM at S3 (old WHO, cumulat from S1)	O		Known DM at S5 (old WHO, cumulat from S1)	BZ	
Known DM (sr+ sr DM drug, S3)	P		Known DM (sr+ sr DM drug, S5)	CA	
DM classif (ADA: OGTT + known DM, S3d)	Q		DM classif (ADA: OGTT + known DM, S5d)	CB	
DM clasiff (WHO: OGTT + known DM, S3d)	R		DM clasiff (WHO: OGTT + known DM, S5d)	CC	
Family hist of diabetes at S1 & S2 (S3d)	S				
Age at DM diagnosis (old WHO)	JN				
First visit where DM was diagnosed (old WHO)	JO				
Last visit before DM was diagnosed (old WHO)	JP				
Follow up time for inc dm from S1 (yrs, old WHO)	JQ				
			Follow up time for inc dm from S3 (yrs, old WHO)	JR	
Year diabetes was diagnosed (S9)	JS				
			Diabetic (screening, S5)	HC	
Q11a Any longstanding illnesses? (S3)	T		Q2.1a Any longstanding illnesses? (S5)	CK	
Q11b Which longstanding illnesses? (S3)	U		Q2.1b Longstanding illness 1 (S5)	CL	

Q11b Which longstanding illnesses? (S3)	V		Q2.1b Longstanding illness 2 (S5)	CM	
Q11b Which longstanding illnesses? (S3)	W		Q2.1b Longstanding illness 3 (S5)	CN	
			Q2.1b Longstanding illness 4 (S5)	CO	
			Q2.1b Longstanding illness 5 (S5)	CP	
			Q2.1b Longstanding illness 6 (S5)	CQ	
Q20a Med presc by doctor last 14 days? (S3)	X		Q2.32a Took presc med last 14days (S5)	CR	
Q20b Which medicine prescribed? (S3)	Y		Q2.32b Medicine taken in last 14 days - 1 (S5)	CS	
Q20b Which medicine prescribed? (S3)	Z		Q2.32b What presc medicine-2 (S5)	CT	
Q20b Which medicine prescribed? (S3)	AA		Q2.32b What presc medicine-3 (S5)	CU	
Q20b Which medicine prescribed? (S3)	AB		Q2.32b What presc medicine-4 (S5)	CV	
			Q2.32b What presc medicine-5 (S5)	CW	
			Q2.32b What presc medicine-6 (S5)	CX	
Drug Class: Anti-hypertensives (S3d)	AC		Drug Class: Anti-hypertensives (S5d)	CY	
Drug Class: diabetic medication(S3d)	AD		Drug Class: Diabetic Medication (S5d)	CZ	(used for exclusion)
			Q2.9b Insulin as treatm for diabetes (S5)	CD	
			Q2.9b Special or diabetic diet? (S5)	CE	
			Drug subclass: Antidepressants (S5d)	DA	
			Drug subclass: Antihistamines (S5d)	DB	
			Drug subclass: Cortico-	DC	

			steroids (S5d)		
			Drug subclass: Insulin (S5d)	DD	
			Drug subclass: Oral antidiabetic (S5d)	DE	
			Drug subclass: Other anti-hypertensives (S5d)	DF	
			Drug subclass: for Rheumatoid arthritis (S5d)	DG	
			Drug subclass: Systemic cortico-steroids (S5d)	DH	
			Drug subclass: Thyroid hormones (S5d)	DI	
Q24b Age periods stopped (S3)	AE		Q3.2b Age periods stopped (S5)	DJ	
Cigarette smoking (never/ex/curr) (S3d)	AI		Cigarette smoking (never/ex/curr) (S5d)	EK	
GHG Anxiety Subscale (S3d)	AJ		GHQ Anxiety subscale (S5d)	EL	
GHQ Chronic group (0/1) (S3d)	AK		GHQ Chronic group (0/1) (S5d)	EM	
GHQ Chronic score (S3d)	AL		GHQ Chronic score (S5d)	EN	
GHQ Depression case (score 4+)(S3d)	AM		GHQ Depression case(score 4+) (S5d)	EO	
GHQ Depression subscale (S3d)	AN		GHQ Depression subscale from GHQ (S5d)	EP	
GHQ score (S3d)	AO		GHQ high score (S5d)	EQ	
GHQ score (S3d)	AP		GHQ score (S5d)	ER	
Q87 Recently been able to concentrate? (S3)	AQ		Q2.33 Recently able to concentrate (S5)	ES	
Q88 Recently lost much sleep? (S3)	AR		Q2.34 Recently lost sleep by worry (S5)	ET	Sleep/ Meditation/ Relaxation

Q89 Recently having restless nights? (S3)	AS		Q2.35 Recently restless nights? (S5)	EU	Sleep/ Meditation/ Relaxation
Q90 Recently been managing to busy? (S3)	AT		Q2.36 Recently kept yourself busy (S5)	EV	Chronic stress
Q91 Recently getting outside as usual? (S3)	AU		Q2.37 Recently getting out (S5)	EW	Chronic stress
Q92 Recently coping as well as others? (S3)	AV		Q2.38 Recently managing well (S5)	EX	Chronic stress
Q93 Recently felt doing things well? (S3)	AW		Q2.39 Recently doing things well? (S5)	EY	Chronic stress
Q94 Recently happy how a task was done? (S3)	AX		Q2.40 Satisf way carried out task/s (S5)	EZ	Chronic stress
Q95 Recently felt warmth close ones? (S3)	AY		Q2.41 Felt warmth and affection (S5)	FA	
Q96 Recently f Easy to mix with others? (S3)	AZ		Q2.42 Felt easy to mix (S5)	FB	
Q97 Recently time chatting to others? (S3)	BA		Q2.43 Spent time chatting (S5)	FC	
Q98 Recently useful role? (S3)	BB		Q2.44 Felt playing a useful role (S5)	FD	
Q99 Recently f Able to make decisions? (S3)	BC		Q2.45 Felt able to decide (S5)	FE	
Q100 Recently f Const Under strain? (S3)	BD		Q2.46 Felt const. under strain (S5)	FF	Chronic stress
Q101 Recently f Unable to fix problems? (S3)	BE		Q2.47 Felt unable to fix probs (S5)	FG	Chronic stress
Q102 Recently found life a struggle? (S3)	BF		Q2.48 Found life a struggle (S5)	FH	Chronic stress
Q103 Recently enjoyed normal activities? (S3)	BG		Q2.49 Enjoyed normal act. (S5)	FI	Chronic stress
Q104 Recently been taking things hard? (S3)	BH		Q2.50 Taking things hard (S5)	FJ	Chronic stress
Q105 Recently been getting scared? (S3)	BI		Q2.51 Getting scared (S5)	FK	Chronic stress
Q106 Recently able to face problems? (S3)	BJ		Q2.52 Able to face problems (S5)	FL	Chronic stress

Q107 Recently found things overwhelming? (S3)	BK		Q2.53 Found things overwhelming (S5)	FM	Chronic stress
Q108 Recently been feeling unhappy? (S3)	BL		Q2.54 Feeling unhappy (S5)	FN	Chronic stress
Q109 Recently been losing confidence? (S3)	BM		Q2.55 Losing confidence (S5)	FO	Chronic stress
Q110 Recently thought self worthless? (S3)	BN		Q2.56 Think yourself worthless (S5)	FP	Chronic stress
Q111 Recently felt life is hopeless? (S3)	BO		Q2.57 Felt life is hopeless (S5)	FQ	Chronic stress
Q112 Recently f. hopeful re the future? (S3)	BP		Q2.58 Felt hopeful re future (S5)	FR	Chronic stress
Q113 Recently been reasonably happy? (S3)	BQ		Q2.59 Been reasonably happy (S5)	FS	Chronic stress
Q114 Recently been feeling nervous? (S3)	BR		Q2.60 Feeling nervous (S5)	FT	Chronic stress
Q115 Recently f. life not worth living ? (S3)	BS		Q2.61 Life not worth living (S5)	FU	Chronic stress
Q116 Recently f. nerves stopped you? (S3)	BT		Q2.62 Nerves stopped you (S5)	FV	Chronic stress
SF36 Gral Mental Health score (S3d)	BU		SF-36 Gral Mental Health score (S5d)	FW	Chronic stress
			Q2.64c Cannot stay asleep (S5)	CF	
			Q2.63 Hrs sleep on avg week night (S5)	CG	
			Q2.64a Trouble falling asleep (S5)	CH	
			Q2.64d Wake as usual but tired (S5)	CI	
			Q2.64b Wake several times (S5)	CJ	

Q51d Hours of mild exercise (S3)	AF				Physical activity pattern
Q51e Hours of mod exercise (S3)	AG				Physical activity pattern
Q51f Hours of vig exercise (S3)	AH				Physical activity pattern
			Mild exerc: Hrs/wk last 4 wks (S5d)	DK	
			Moderate exerc: Total Hours (S5d)	DL	
			Physical Activ Level WHO, hrs+frq, Sabia (S5d)	DM	
			Physical Activ Level, hrs, Sabia (S5d)	DN	
			Mild activ hrs/wk, Sabia (S5d)	DO	
			Mild activ MET-hrs/wk, Sabia (S5d)	DP	Physical activity pattern
			Mild activ occns/wk, Sabia (S5d)	DQ	
			Moderate activ MET-hrs/wk, Sabia (S5d)	DR	Physical activity pattern
			Moderate activ hrs/wk, Sabia (S5d)	DS	
			Moderate activ occns/wk, Sabia (S5d)	DT	
			Vigorous activ MET-hrs/wk, Sabia (S5d)	DU	Physical activity pattern
			Vigorous activ hrs/wk, Sabia (S5d)	DV	
			Vigorous activ occns/wk, Sabia (S5d)	DW	

			Q4.2a Sport activity-oth 1-1 (S5)	DX	
			Q4.2a Sport activ- oth1: occasions (S5)	DY	
			Q4.2a Sport activ-oth 1: total hours (S5)	DZ	
			Q4.2a Sport activity- oth 2-1 (S5)	EA	
			Q4.2a Sport activ- oth 2: occasions (S5)	EB	
			Q4.2a Sport activ- oth 2: hrs (S5)	EC	
			Q4.2a Swim- Occasions (S5)	ED	
			Q4.2a Swim- Total hours (S5)	EE	
			Vigorous exerc: Total hours S5d)	EF	
			Q4.3 Vig phys activ: total hours/wk (S5)	EG	
			Q4.1a Avg mins walked on week days (S5)	EH	
			Q4.1a Avg mins walked on weekend days (S5)	EI	
			Total METs for walking (S5d)	EJ	
Age at clinical screening (S3d)	FX		Age at clinical screening (S5d)	HD	Age
Body mass index (S3)	FY		Body mass index (S5)	HE	BMI
Height in cm (S3)	FZ		Height cms (S5)	HF	
Waist L3-4 (S3)	GA		Waist largest (S5)	HG	
Waist smallest (S3)	GB		Waist smallest (S5)	HH	
Weight in kg (S3)	GC		Weight in kgs (S5)	HI	
Adiponectin (S3)	GD		Adiponectin (S5)	HJ	Inflammation

CRP inflammatory marker, mg/L (S3)	GE		CRP inflamm marker, mg/L (S5)	HK	Inflammation
Fibrinogen (S3)	GF		Fibrinogen (S5)	HL	Inflammation
IL1 receptor antagonist (ng/mL) (S3)	GG		IL1 receptor antagonist (ng/mL) (S5)	HM	Inflammation
IL6 inflammatory marker, pg/mL (S3)	GH		IL6 inflamm marker, pg/mL (S5)	HN	Inflammation
Time of last food (S3)	GK		Food 1st (S5)	HO	
Time of blood sample (S3)	GI		Time 1st blood sample taken (S5)	HP	
			Time 2nd blood sample taken (S5)	HQ	
Time glucose taken (S3)	GJ		Time of glucose drink (S5)	HR	
Hours of fasting (S3)	GL		Hours of fasting (S5-D)	HS	
Participant fasted (S3d)	GM		Did participant fast? (S5)	HT	(used for exclusion)
Glucose classification: 2 hrs (S3)	GN		Glucose classification: 2 hrs (S5)	HU	
Glucose classification: fasting/1st (S3)	GO		Glucose classification: fasting/1st (S5)	HV	
Glucose 2hrs (mmol/L, S3)	GP		Glucose 2hrs (mmol/L, S5)	HW	
Glucose: fasting/1st (mmol/L,S3)	GQ		Glucose: fasting/1st (mmol/L, S5)	HX	Fasting glucose
HOMA- Beta Cell function (S3)	GR		HOMA- Beta Cell function (S5)	HY	Beta cell function
HOMA- Insulin Resistance (S3)	GS		HOMA- Insulin Resistance (S5)	HZ	
HOMA- Insulin Sensitivity (S3)	GT		HOMA- Insulin Sensitivity (S5)	IA	Insulin sensitivity
Insulin 2 hrs (uIU/mL, S3)	GU		Insulin 2 hrs (uIU/mL, S5)	IB	
Insulin: fasting/1st (uIU/mL, S3)	GV		Insulin: fasting/1st (uIU/mL, S5)	IC	Fasting Insulin
Cholesterol - Total (S3)	GW		Blood cholesterol (S5)	ID	Tissue damage
Cholesterol - HDL (S3)	GX		High density lipoprotein (S5)	IE	Tissue damage

Cholesterol - LDL (S3)	GY		LDL cholesterol, mm (S5)	IF	Tissue damage
Triglyceride (S3)	GZ		Triglyceride (S5)	IG	Tissue damage
Diastolic BP- final averaged (S3-d)	HA		Diastolic BP- final averaged (S5)	IH	Tissue damage
Systolic BP - final averaged (S3-d)	HB		Systolic BP- final averaged (S5)	II	Tissue damage
Avg beer, lager or cider (FFQ3)	IJ		Avg beer, lager or cider (FFQ5)	IO	Food quality
Avg liqueurs (FFQ3)	IK		Avg liqueurs (FFQ5)	IP	Food quality
Avg port, cherry, vermouth (FFQ3)	IL		Avg port, cherry, vermouth (FFQ5)	IQ	Food quality
Avg spirits (FFQ3)	IM		Avg spirits (FFQ5)	IR	Food quality
Avg wine (FFQ3)	IN		Avg wine (FFQ5)	IS	Food quality
Energy value in kJ (FFQ3-F)	IT		Energy value in kJ (FFQ5-F)	JD	
Energy value in kcal (FFQ3-F)	IU		Energy value in kcal (FFQ5-F)	JE	Food intake pattern
Fibre (FFQ3-F)	IV		Fibre (FFQ5-F)	JF	Food quality
Sat fa exc branch g/100g fa (FFQ3-F)	IW		Sat fa exc branch g/100g fa (FFQ5-F)	JG	Food quality
Monunsat fa g/100g fa (FFQ3-F)	IX		Monunsat fa g/100g fa (FFQ5-F)	JH	Food quality
Energy value in kJ (FFQ3-S)	IY		Energy value in kJ (FFQ5-S)	JI	
Energy value in kcal (FFQ3-S)	IZ		Energy value in kcal (FFQ5-S)	JJ	
Monunsat fa g/100g fa (FFQ3-S)	JC		Monunsat fa g/100g fa (FFQ5-S)	JM	

Fibre (FFQ7-S)	JK		Fibre (FFQ5-S)	JA	
Sat fa exc branch g/100g fa (FFQ7-S)	JL		Sat fa exc branch g/100g fa (FFQ5-S)	JB	
ILLNESS_S3_FLAG_I	JT		ILLNESS_S5_FLAG_I	KA	
ILLNESS_S3_FLAG_II	JU		ILLNESS_S5_FLAG_II	KB	
ILLNESS_S3_FLAG_III	JV		ILLNESS_S5_FLAG_III	KC	
			ILLNESS_S5_FLAG_IV	KD	
			ILLNESS_S5_FLAG_V	KE	
			ILLNESS_S5_FLAG_VI	KF	
MEDICINE_S3_I	JW		MEDICINE_S5_I	KG	(used for exclusion)
MEDICINE_S3_II	JX		MEDICINE_S5_II	KH	(used for exclusion)
MEDICINE_S3_III	JY		MEDICINE_S5_III	KI	(used for exclusion)
MEDICINE_S3_IV	JZ		MEDICINE_S5_IV	KJ	(used for exclusion)
			MEDICINE_S5_V	KK	(used for exclusion)
			MEDICINE_S5_VI	KL	(used for exclusion)
			MEDICINE_S5_VII	KM	(used for exclusion)
			MEDICINE_S5_VIII	KN	(used for exclusion)
			MEDICINE_S5_IX	KO	(used for exclusion)
			MEDICINE_S5_X	KP	(used for exclusion)

The following (unnumbered) tables show some basic descriptives of the transferred Whitehall II dataset.

Age related variables with selection of subjects participated both in S3 & S5.

Variable	N	mean	std	min	max
Year of birth	7666	41.8	6.1	30.0	52.0
Age at participation date S3	7666	50.2	6.1	39.6	64.1
Age at participation date S5	7666	56.0	6.0	44.8	69.2
Age at quest completion S3	7562	50.2	6.1	39.6	64.1
Age at quest completion S5	7629	56.0	6.0	44.8	69.2
Age at clinical screening S3	7192	50.0	6.0	39.6	63.3
Age at clinical screening S5	6442	55.8	6.0	44.8	69.0

Frequency table of participated subjects in S3 by gender, ethnicity, & smoking status

Obs	Sex	Ethnicity	Cigarette smoking	Frequency Count	Percent
1	MALE	white	.	572	.
2	MALE	white	NEVER-SMOKER	2239	28.6
3	MALE	white	EX-SMOKER	2150	27.5
4	MALE	white	CURRENT SMOKER	641	8.2
5	MALE	non-white	.	70	.
6	MALE	non-white	NEVER-SMOKER	179	2.3
7	MALE	non-white	EX-SMOKER	132	1.7
8	MALE	non-white	CURRENT SMOKER	74	0.9
9	FEMALE	white	.	261	.
10	FEMALE	white	NEVER-SMOKER	975	12.5
11	FEMALE	white	EX-SMOKER	709	9.1
12	FEMALE	white	CURRENT SMOKER	408	5.2
13	FEMALE	non-white	.	95	.
14	FEMALE	non-white	NEVER-SMOKER	254	3.2
15	FEMALE	non-white	EX-SMOKER	33	0.4
16	FEMALE	non-white	CURRENT SMOKER	23	0.3

Frequency table of participated subjects in S5 by gender, ethnicity, & smoking status

Obs	Sex	Ethnicity	Cigarette smoking	Frequency Count	Percent
1	MALE	white	.	382	.
2	MALE	white	NEVER-SMOKER	2198	30.5
3	MALE	white	EX-SMOKER	2090	29.0
4	MALE	white	CURRENT SMOKER	446	6.2
5	MALE	non-white	.	34	.
6	MALE	non-white	NEVER-SMOKER	171	2.4
7	MALE	non-white	EX-SMOKER	114	1.6
8	MALE	non-white	CURRENT SMOKER	38	0.5

Obs	Sex	Ethnicity	Cigarette smoking	Frequency Count	Percent
9	FEMALE	white	.	198	.
10	FEMALE	white	NEVER-SMOKER	925	12.9
11	FEMALE	white	EX-SMOKER	675	9.4
12	FEMALE	white	CURRENT SMOKER	272	3.8
13	FEMALE	non-white	.	58	.
14	FEMALE	non-white	NEVER-SMOKER	233	3.2
15	FEMALE	non-white	EX-SMOKER	25	0.3
16	FEMALE	non-white	CURRENT SMOKER	11	0.2

Frequency table of subjects with diabetic medication in S3.

Obs	Drug Class: Diabetic medication	Frequency Count	Percent
1	.	8759	.
2	YES	56	100

Frequency table of subjects with insulin as diabetic treatment in S5.

Obs	Insulin as treatm for diabetes	Frequency Count	Percent
1	.	7700	.
2	Yes	44	25.9
3	No	126	74.1

Frequency table of subjects with diabetic medication in S5.

Obs	Drug Class: Diabetic medication	Frequency Count	Percent
1	.	68	.
2	NO	7669	98.3
3	YES	133	1.7

Frequency table of subjects with insulin as drug class in S5

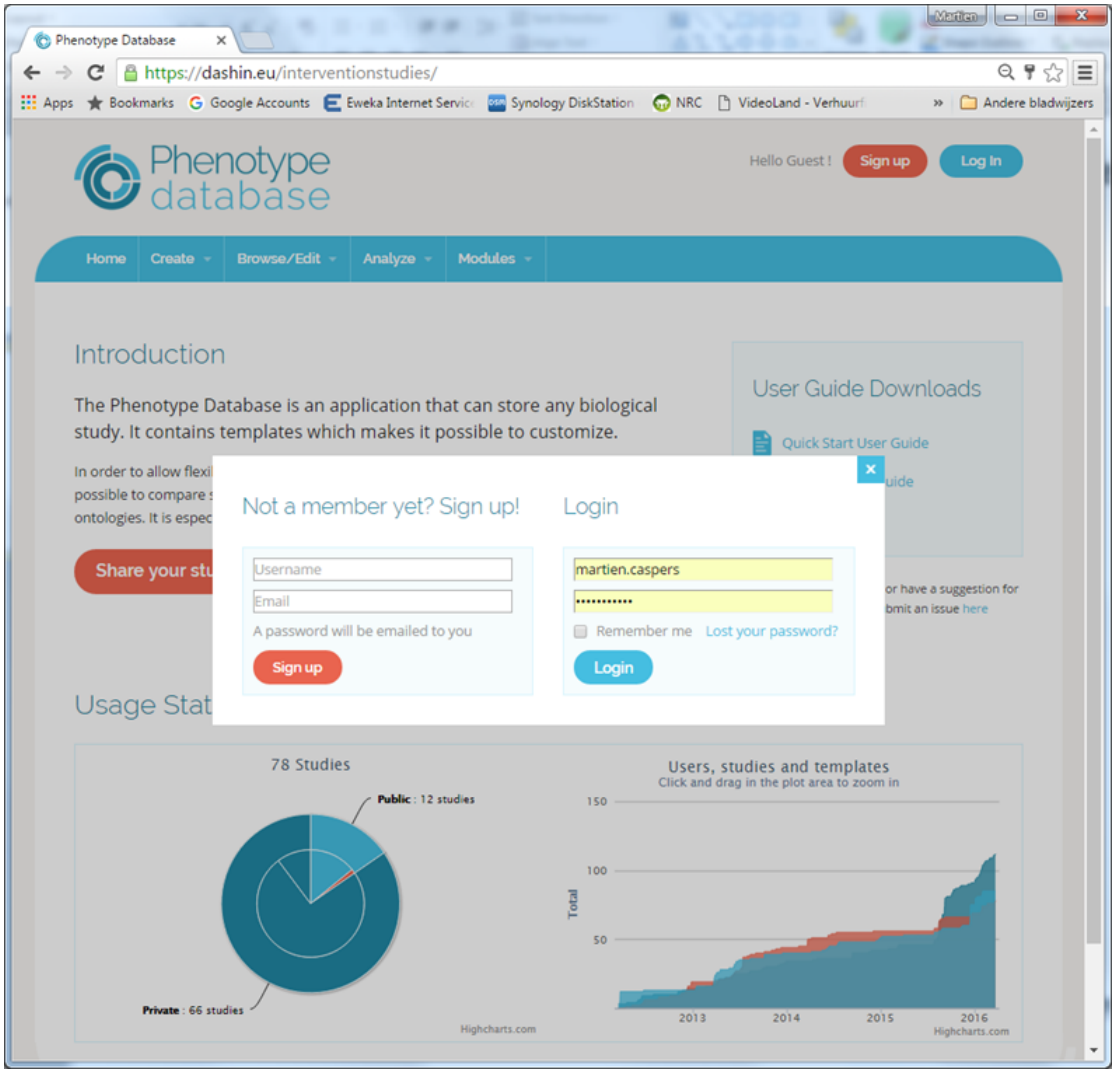
Obs	Drug subclass: Insulin	Frequency Count	Percent
1	.	68	.
2	NO	7759	99.4
3	YES	43	0.6

4.2 Annex 2: Nutritional Phenotype Database (dbNP) Access & Data-Download

How to access the Nutritional Phenotype database (developed in NutriTech)?

Website: http://studies.dbnp.org/

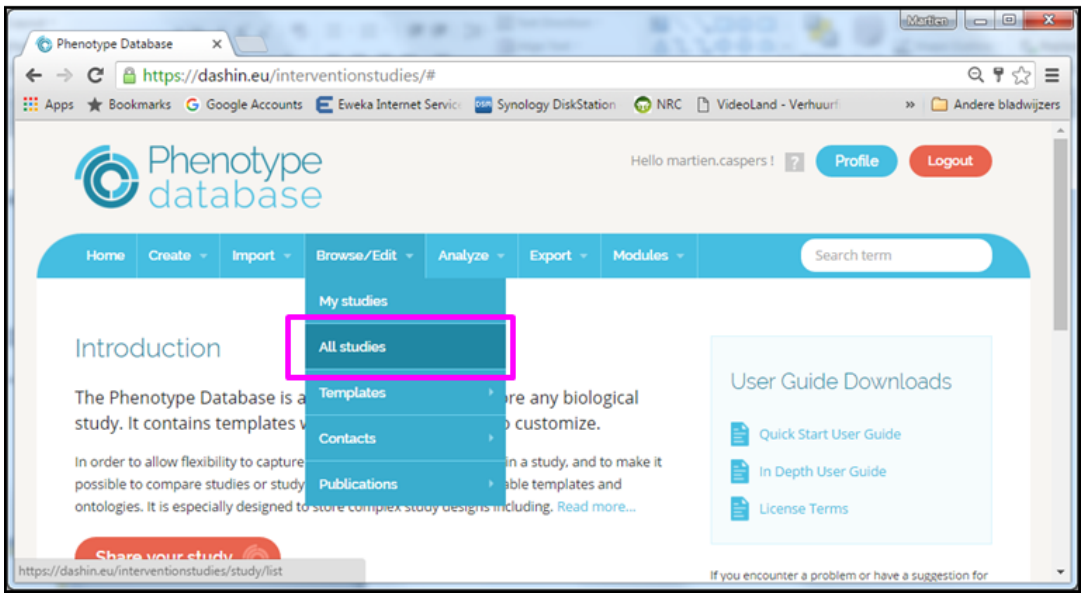
- Sign up: Username: firstname.lastname
  email: yours@nnn.nn →after sign up, a password will be emailed to you.
- Login: Username: firstname.lastname
  Password: xxxxxx



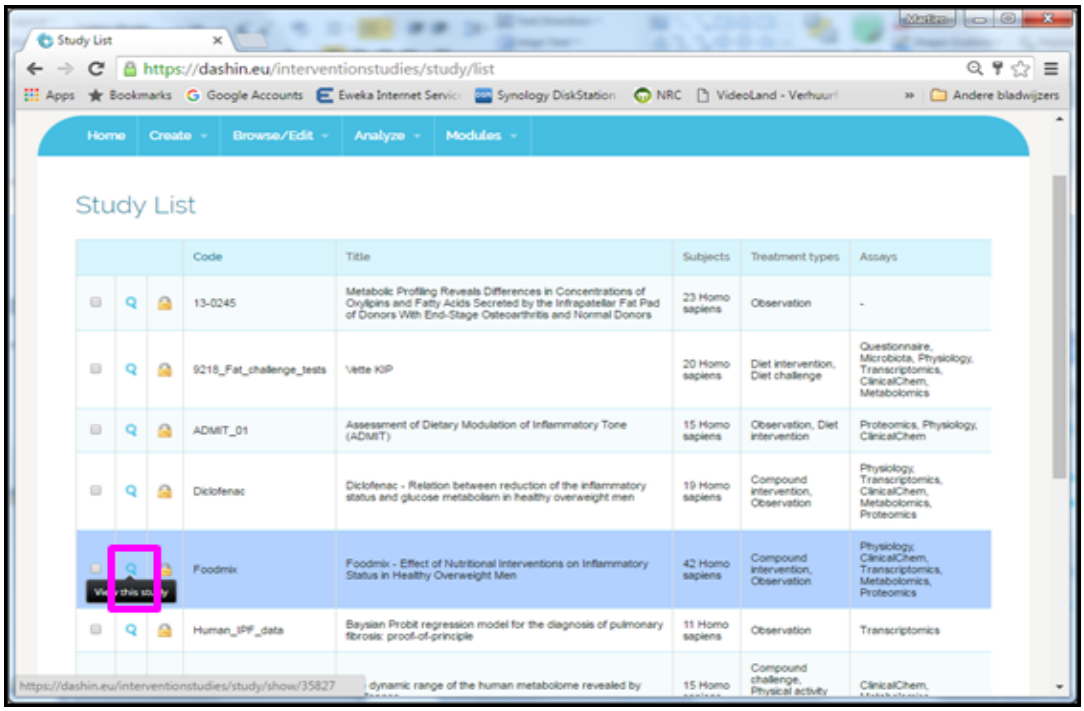
Search and view a study (steps 1 to 5):

1) Browse/Edit (upper blue menu-bar)
2) View studies or My Studies or All studies (drop down menu):

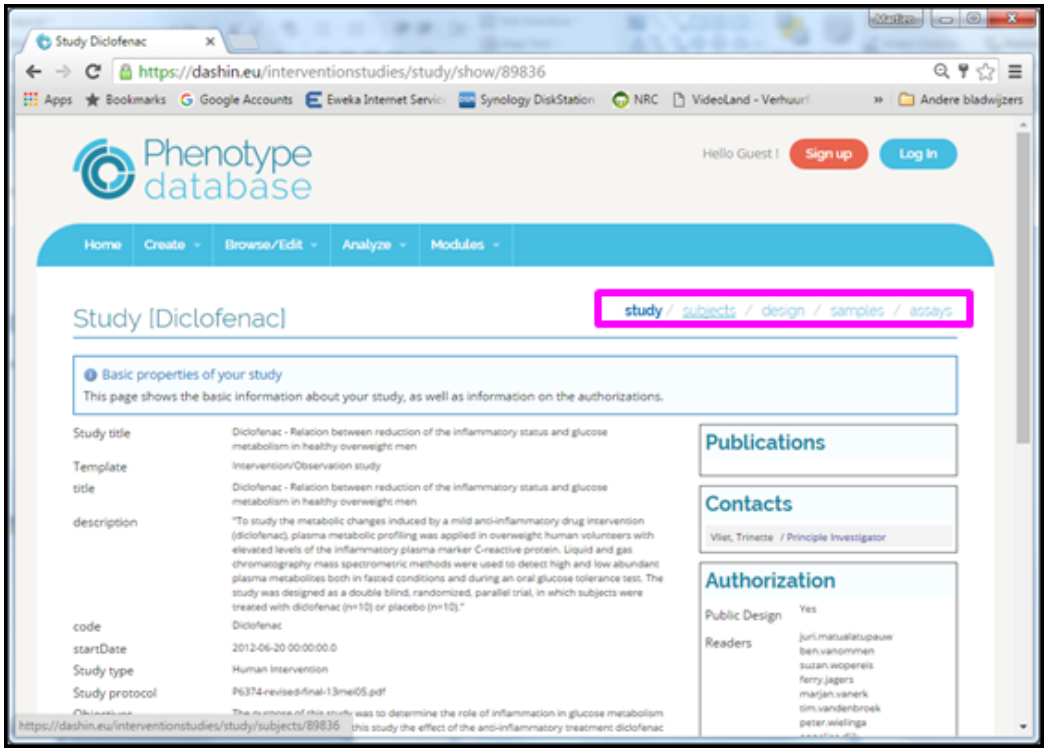
= Menu/Click/Scroll location



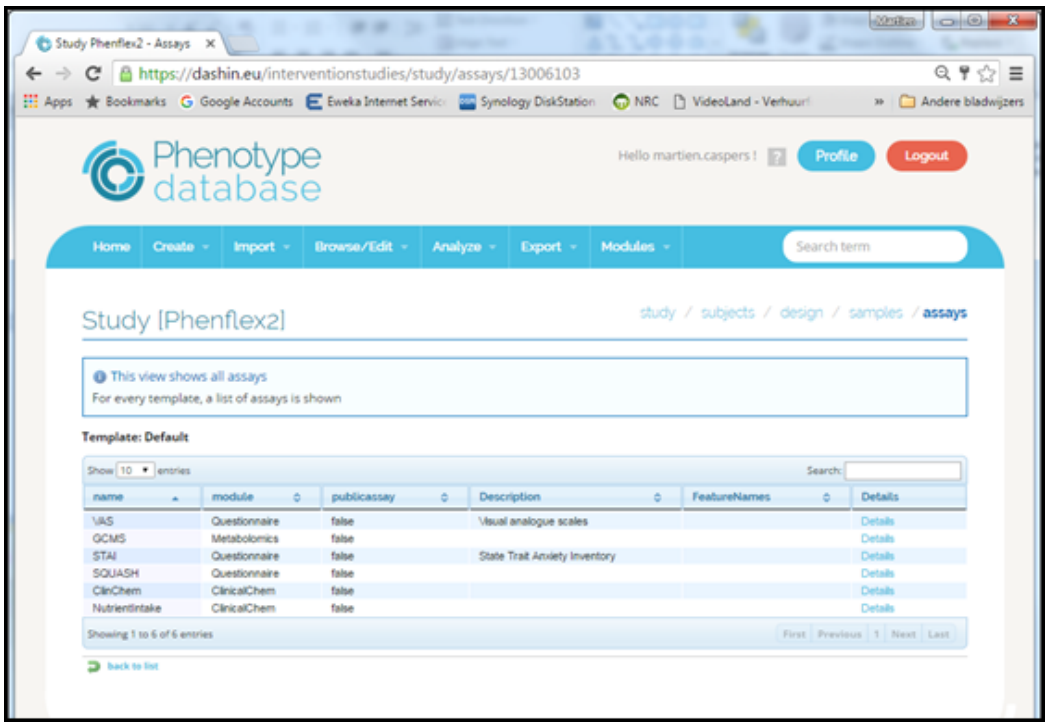
3) Select your Study choice (from Study List):



4) View: Study, Subjects, Design, Samples, Assays (below menu bar, right)

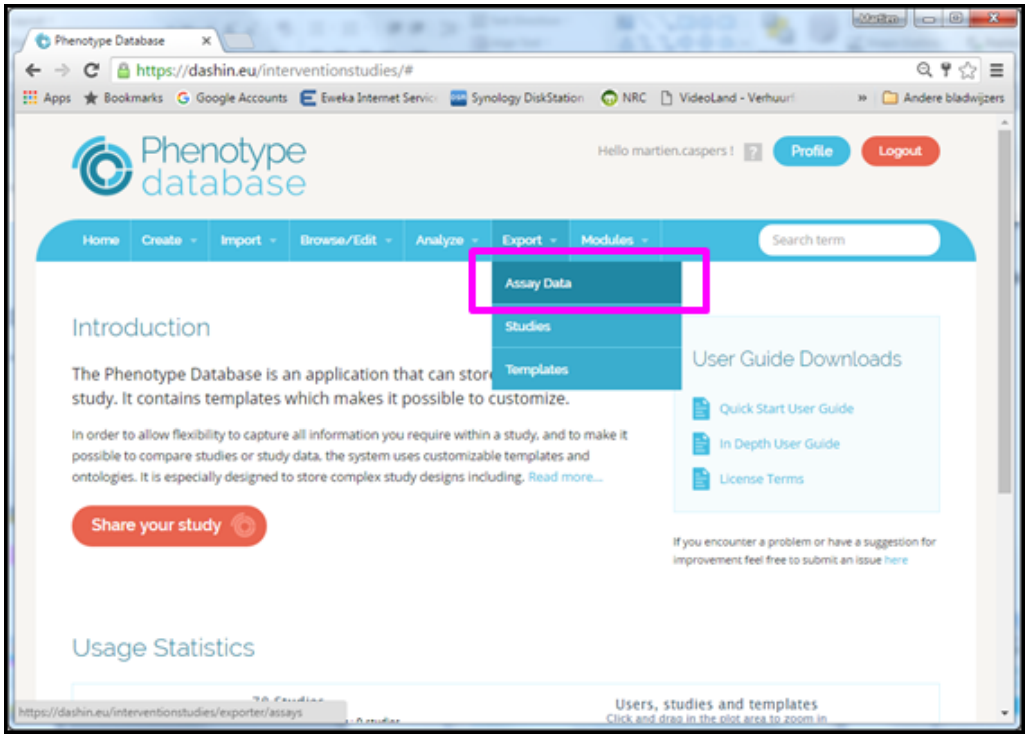


5) Identify data sets: Assays

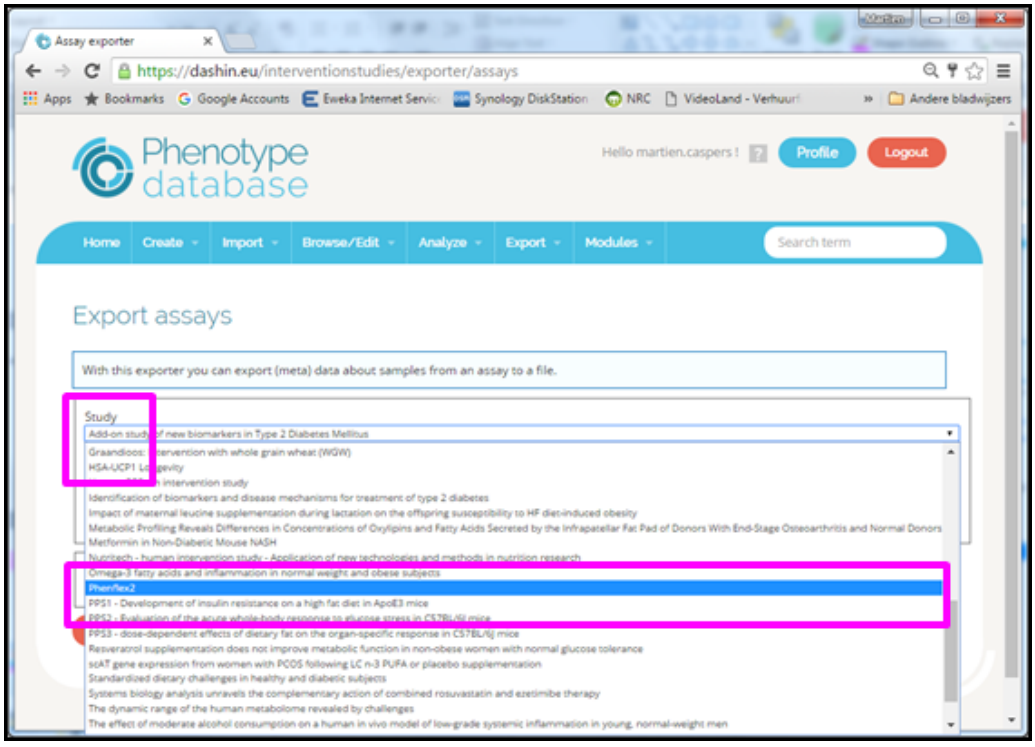


Download study data (you should be logged in) (steps 1 to 4):

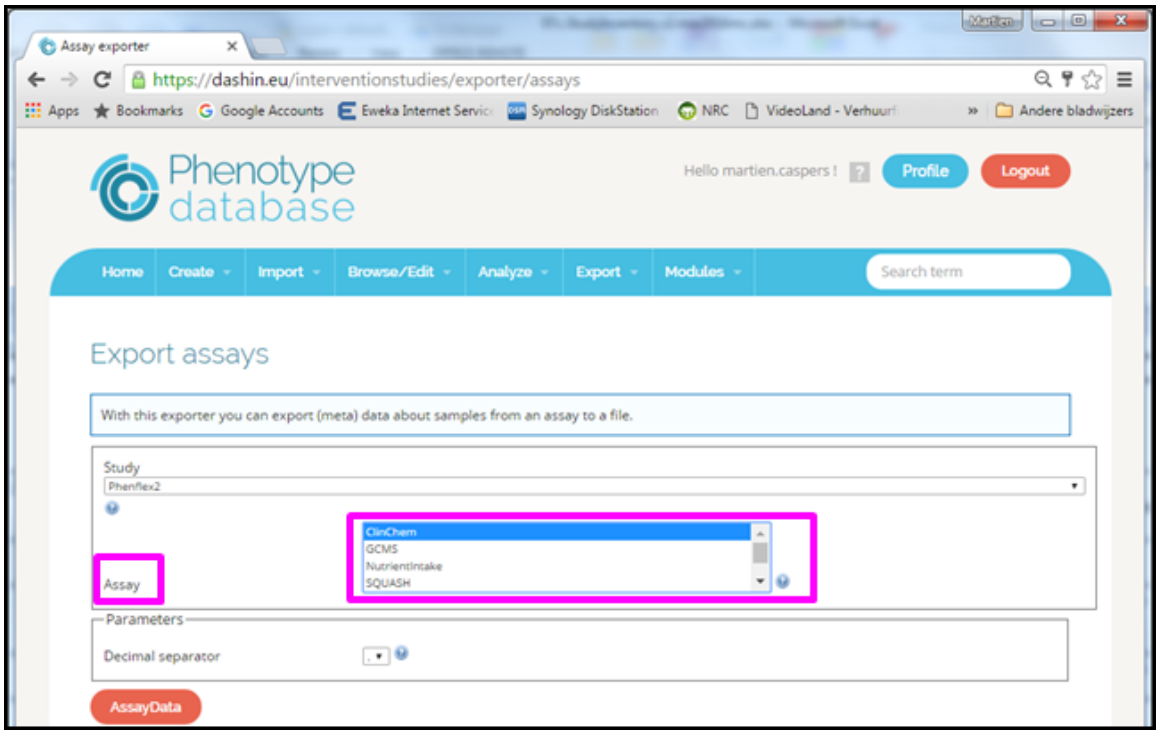
1) Export (upper blue menu-bar)
2) Assay data (drop down menu)



3) Select your study (from “Study” drop down menu, scroll and click)



4) Select an assay (from “Assay” drop down, scroll & click) and click AssayData → download starts



Layout of Exported Data: