This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 774548.

EC Framework Programme for Research and Innovation

Horizon 2020

H2020-SFS-2017-2-RIA-774548-STOP:

Science & Technology in childhood Obesity Policy

Science and Technology in childhood Obesity Policy

Science & Technology in childhood Obesity Policy

Start date of project: 1st June 2018 Duration: 48 months

D1.2: Data Management Plan

Author(s): Ella Hattey, Imperial College London

Version: Final

Preparation date: 07/04/2020

Dissemination Level

PU	Public	☒
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Science and Technology in childhood Obesity Policy

Abbreviation	Definition
NCD-RisC	Non communicable disease risk factor collaboration
SES	Socio-economic status
EFSA	European Food Standards Agency
SUSFOOD	Sustainable food consumption and production
Food-EPI	Healthy Food Environment Policy Index
RCT	Randomised control trial

Overview of data sets used in STOP

Data Set	Study	Population	Sample size	Access	Data protection	Ethical approval
NCD-RisC & SES analysis	Openly accessible data sets (approx. 3000 studies)	Community, national and subnational studies across Europe	N/A (multiple studies which vary in sample size)	Openly accessible data sets, downloadable online	We use fully anonymised sources many of which are downloadable online (for example Health Survey for England etc).	We only use population mean/prevalence in our Bayesian model.
WP 3 Data Set(s)	Rhea Inma Piccoli+ Environage ALSPAC Generation 21 Epiteen Timisoara Zagreb Ljubljana HELIX Subcohort WP8 trial sites	Greece Spain Italy Belgium UK Portugal Portugal Romania Croatia Slovenia EU Countries Sweden, Spain, Romania	Rhea – 1500 Inma – 1500 Piccoli+ - 3328 Environage – 1300 ALSPAC – 14062 Generation 21 – 8674 Epiteen – 2942 Timisoara – 200 Zagreb – 895 Ljubljana – 1200 HELIX Subcohort (includes INMA, Rhea) – 1200 WP8 trial sites – 300 families	Data to third parties available upon request, but not in an open repository. All project partners can freely access the data.	Data Transfer Agreements between the cohorts and Imperial College are in place which outlines the use and transfer of data. The data used in STOP is anonymised and does not include any biometric or genetic indicators that would allow for identification of individuals. In this regard the used dataset is not subject to GDPR.	Ethics approval is the responsibility of each cohort and has been granted by each institution's ethical review process.

EFSA dietary survey micro-data – WP4	European Food Safety Authority	Finland, France, Italy, Spain, and the United Kingdom Individuals' food consumption data and calories, fat and carbohydrates intakes for children and adolescents.	Finland – 2022 France – 1444 Italy – 454 Spain – 86 United Kingdom - 3664	Project partners	Restricted. Need EFSA's agreement.	No ethical approval is necessary.
ERA-Net SUSFOOD Consortium SUSDIET elasticities- WP4	ERANET SUSDIET Project (Akaichi, F. et al, 2017).	Finland, France, Italy, Spain, and the United Kingdom Price elasticities for 14 food product categories, calculated from home scanner data	One price elasticity by food group and country. Calculated for a representative household	From SUSDIET's contributors	No protection	No ethical approval is necessary.
Home scanner dataset- WP4	Home scanner data from KANTAR UK	Different food categories from the United Kingdom (2015, 2017 and 2018), Spain (2017) and France (2015, 2016, 2017).	N/A	The access will be contractually restricted to a set of individuals from ICL and INRAE	INRAE and ICL have appropriate technical and organisational measures in place to protect any data against unauthorised or unlawful data transfer, processing and against accidental loss, destruction or damage.	No ethical approval is necessary for this kind of data
WP4 Food-EPI	Implementing the Food-EPI policy index	NGO experts, academic experts	Different from country to country, between 20 and 60	Project partners	Compliant with GDPR	Arranged by each participating country

WP5 data set	Behavioural experiment	TBC, but likely school children in Slovenia	TBC	TBC	All data collected will be stored on a secure server and only accessible to the researchers involved.	Ethics approval will be sought through the beneficiary's institution.
Healthy Lifestyle (HLS)	SLOfit	Slovenia	30,000	Data to third parties available upon request, but not in an open repository. All project partners can freely access the data.	The data used in STOP is anonymised and does not include any biometric or genetic indicators that would allow for identification of individuals. In this regard the used dataset is not subject to GDPR.	Ethical approval for the SLOfit study was obtained from the National Medical Ethics Committee of the Republic of Slovenia, ID 102/03/15
RCT Dataset, (WP8):	Randomised control trial	Children 2-6 years old with overweight and obesity referred to the study from child health care centres and paediatric clinics in Stockholm, Sweden, Timisoara, Romania and in Mallorca, Spain.	n=300 (100 children from each study site)	Agreements to be signed with other WPs with which collaboration will take place.	All collected data is stored at the University's server at each study site. Only researcher involved in the study has access to collected data.	Ethics approval was obtained from: the Ethics Committee of Scientific Research in University of Medicine and Pharmacy "Victor Babes", Timisoara, Romania, October 31st, 2018 (25/31.10.2018), the Balearic Islands Ethics Committee, Mallorca, Spain, February 13th, 2019 (IB 3814/18

						PI), and the Research Ethics Committee, Stockholm, Sweden, December 11th, 2018 (2018/2082-31/1) with amendment January 16th 2020 (2019-05593). Written informed consent is obtained from all parents or caregivers.

Data Management Plan

The document below was created using DMP Online. DMP Online helps to create, review, and share data management plans that meet institutional and funder requirements. It is provided by the Digital Curation Centre (DCC).

STOP — Science & Technology in childhood Obesity Policy, H2020-GA774548 - Initial DMP

1. Data summary

Provide a summary of the data addressing the following issues:

- **State the purpose of the data collection/generation**
- **Explain the relation to the objectives of the project**
- **Specify the types and formats of data generated/collected**
- **Specify if existing data is being re-used (if any)**
- **Specify the origin of the data**
- **State the expected size of the data (if known)**
- **Outline the data utility: to whom will it be useful**

NCD-RisC & SES analysis- WP2

Name of the Data Set	NCD-RisC & SES analysis
State the purpose of the data collection/generation	This WP will generate comparable estimates of measures of obesity, and more broadly anthropometry, for countries in Europe. The process involves identifying and pooling existing population-based studies that have measured height and weight and using a statistical model to make estimates by age, gender, country and year.
Explain the relation to the objectives of the project	STOP will generate estimates of measures of obesity and more broadly anthropometry, nationally as well as by rural and urban place of residence and SES.
Specify the types and formats of data generated/collected	As below, this WP will use a database on cardiometabolic risk factors collated by the Non-Communicable Disease Risk Factor Collaboration (NCD-RisC). In bringing together this database, we excluded all data sources that were solely based on self-reported weight and height without a measurement component because these data are subject to biases that vary by geography, time, age, sex and socioeconomic characteristics. Due to these variations, approaches to correcting self-reported data leave residual bias. We also excluded data sources on population subgroups whose anthropometric status may differ systematically from the general population, including: • studies that had included or excluded people based on their health status or cardiovascular risk; • studies whose participants were only ethnic minorities; • specific educational, occupational, or socioeconomic subgroups, with the exception noted below; • those recruited through health facilities, with the exception noted below; and women aged 15-19 years in surveys which sampled only ever-married women or measured height and weight only among mothers.
Specify if existing data is being re-used (if any)	In NCD-RisC database, anonymised individual record data from sources included in NCD-RisC were reanalysed by core NCD-RisC researchers or by data holders according to a common protocol. We dropped participants with implausible BMI levels or with implausible height or weight values, which are all age dependent. All analyses incorporated appropriate sample weights and complex survey design, when applicable, in calculating age-sex-specific means and prevalence. To ensure summaries were prepared according to the study protocol, the NCD-RisC core researchers provided computer code to NCD-RisC members who requested assistance. All submitted data were checked by at least two independent NCD-RisC core researchers. Questions and clarifications were discussed with NCD-RisC members and resolved before data were incorporated in the database. Finally, we incorporated all nationally representative data from sources that were identified but not accessed via the above routes, by extracting summary statistics from published reports. Data were also extracted for four sub-national STEPS surveys, one Countrywide Integrated Non-communicable Diseases Intervention (CINDI) survey, and six sites of the WHO Multinational MONItoring of trends and determinants in CArdiovascular disease (MONICA) project that were not deposited in the MONICA Data Centre. Data were extracted from published reports only when reported by sex and in age groups no wider than 20 years. We also used data from a previous global-data pooling study when such data had not been accessed through the routes described. All NCD-RisC members are asked periodically to review the list of sources from their country, to suggest additional sources not in the database, and to verify that the included data meet the inclusion criteria listed below and are not duplicates. The NCD-RisC database is continuously updated through this contact with NCD-RisC members and all the above routes.
Specify the origin of the data	This WP will use a database on cardiometabolic risk factors collated by the Non-Communicable Disease Risk Factor Collaboration (NCD-RisC). NCD-RisC is a worldwide network of health researchers and practitioners whose aim is to document systematically the worldwide trends and variations in NCD risk factors. The database was collated through multiple routes for identifying and accessing data. We accessed publicly available population-based measurement surveys (e.g., Demographic and Health Surveys (DHS), Global School-based Student Health Surveys (GSHS), the European Health Interview and Health Examination Surveys (EHIS and EHES) and those available via the Inter-University Consortium for Political and Social Research (ICPSR). We requested, via the World Health Organization (WHO) and its regional and country offices, help with identification and access to population-based surveys from ministries of health and other national health and statistical agencies. Requests were also sent via the World Heart Federation to its national partners. We made similar requests to the co-authors of an earlier pooled analysis of cardiometabolic risk factors and invited them to reanalyse data from their studies and join NCD-RisC. Finally, to identify major sources not accessed through the above routes, we searched and reviewed published studies, and invited all eligible studies to join NCD-RisC.
State the expected size of the data (if known)	Data will be analysed for every country in Europe and for every year, since 1975 nationally, probably 1985 by rural and urban place of residence, and around the same time or bit later for SES. STOP will generate estimates of measures of obesity and more broadly anthropometry, nationally as well as by rural and urban place of residence and SES.
Outline the data utility: to whom will it be useful	Other STOP WPs

WP3 Data Set(s)

Name of the Data Set	WP 3 Data Set(s)
State the purpose of the data collection/generation	The purpose of the Data collection/generation is to measure the impact of the obesogenic behaviours of children and their families, to understand their health consequences over the life course, and how this vary by SES. It will also help to identify the weakest links in the causal pathways that allow effective interventions
Explain the relation to the objectives of the project	The objectives of the project are to show how childhood obesity has spread and its long –term consequences on health, social, and economic outcomes. The primary focus of the project is on the cumulative impacts of multiple and synergistic exposures in vulnerable and socially disadvantaged children and their families. **Main points** • The project aims to expand the multi-disciplinary evidence base upon which effective and sustainable policies can be built to prevent and manage childhood obesity. • The project aims at creating the conditions for evidence to translate into policy and then for policy to translate into impacts on the ground. • The project aims to generate new evidence using the full range of investigative approaches that have the potential to contribute to the development and implementation of effective childhood obesity policies, including laboratory assessment of bio markers, randomised experimentation of novel interventions, quasi-experiments to test behavioural incentives and new technologies, secondary statistical analyses of large-scale datasets, simulation modelling of policy impacts on health, social and economic outcomes at the population level, as well as analyses based on combinations of the above approaches aimed at cross-validating evidence and policy recommendations. • The Project will generate scientifically sound, policy relevant evidence on the factor that has been contributed to the spread of childhood obesity in European countries and on the effect of alternatives technological and organisational solution and policy options available to address the problems. Another aim is to bring together major health and food sector actors, including scientist, health professional, government policy maker international organisation and many more through which policy-relevant evidence can be generated, made available and used in the design and implementation of effective of childhood obesity problem in EU, national and local level.
Specify the types and formats of data generated/collected	WP3 will rely mainly on re-analyses of existing data (questionnaires, “omics” measurements in biological samples, clinical measurements), supplemented by new gut hormone measurements and geographic information system (GIS) linkages to generate new exposure data. New derived variables will also be generated from existing data including ultra-processed food scores and dietary scores derived from nuclear magnetic resonance spectroscopy (NMR) data. A summary of the different analyses of data is given in table 1 **New data** will be generated as follows: • 1000 measurements of gut hormones in blood (ghrelin, PYY and GLP-1) will be measured by radioimmunoassay, selected from samples collected in the Rhea, Inma, Generation 21 and Timisoara cohorts. Data will be processed into standard formats (e.g. .csv files) • GIS data will be collected from relevant authorities, databases and existing exposure models and linked to geocoded addresses available in cohorts selected from Environage, Generation 21, Timisoara, Zagreb and Ljubliana. Data will be processed into standard formats (e.g. .csv files) **New derived variables (to be processed into .csv files):** • NMR derived dietary scores in the HELIX exposome cohort Questionnaire derived ultra-processed food scores from cohorts selected from ALSPAC, Generation 21, Inma, Rhea and HELIX, Timisoara and Zagreb cohorts
Specify if existing data is being re-used (if any)	Existing data includes: • questionnaire and clinical examination data (all cohorts) • “omics” data from ALSPAC, Generation 21, Inma, Rhea, Envionage, Piccolipiu, and HELIX GIS derived exposure data from HELIX, INMA, RHEA, Piccolipiu, ALSPAC cohorts
Specify the origin of the data	Existing data and samples (from which gut hormones will be measured) will come from cohorts based in Greece, Spain, Italy, Belgium, Portugal, Romania, Croatia, Slovenia, Lithuania, Norway, France & United Kingdom. GIS data will be derived from multiple sources including local municipal authorities, European organisations (European Environment Agency), commercial databases, (e.g. NAVTEQ), public databases (e.g. urban atlas, open street maps) and National statistics offices. These data will be linked to address information collected in the cohorts.
State the expected size of the data (if known)	< 500 GB
Outline the data utility: to whom will it be useful	The data will primarily be useful to the research community

EFSA dietary survey micro-data – WP4

Name of the Data Set	EFSA dietary survey micro-data – WP4
State the purpose of the data collection/generation	Food consumption data were collected by EFSA for assessing how exposed people are to potential risks in the food chain
Explain the relation to the objectives of the project	The EFSA micro dataset (food composition and consumption) will be used to assess the impact of fiscal policy on product consumptions and intakes.
Specify the types and formats of data generated/collected	Csv file
Specify if existing data is being re-used (if any)	yes
Specify the origin of the data	European Food Safety Authority (EFSA)
State the expected size of the data (if known)	Unknown
Outline the data utility: to whom will it be useful	WP4 and WP6 partners

ERA-Net SUSFOOD Consortium SUSDIET elasticities- WP4

Name of the Data Set	ERA-Net SUSFOOD Consortium SUSDIET elasticities- WP4
State the purpose of the data collection/generation	Assess household's price sensitivity for 20 food groups (Fruits; Grains; Starchy roots; Soft drinks; Sugar, desserts; Tea, coffee, water; Vegetables; Plant based fats; Animal fats; Beef/ lamb; Cheese; Fish and seafood; Milk, dairy products; Pork; Poultry, other meat"; Processed, cooked meat; Alcoholic beverages; Composite dishes; Snacks and other foods products) for Finland, France, Italy, Spain, Sweden and UK.
Explain the relation to the objectives of the project	We will use these elasticities to compare the impacts of sugar-sweetened beverages tax, fruit and vegetable subsidy and the combination of the two policies for Finland, France, Italy, Spain, Sweden.
Specify the types and formats of data generated/collected	Csv file
Specify if existing data is being re-used (if any)	Yes
	ERA-Net SUSFOOD Consortium SUSDIET (2014-2017) - Coordinator: Soler, Louis-Georges (INRA)
State the expected size of the data (if known)	455 Ko
Outline the data utility: to whom will it be useful	INRA but can be open to other partners

Home scanner dataset- WP4

Name of the Data Set	Home scanner dataset- WP4
State the purpose of the data collection/generation	Analyse food purchases and product composition The home scanner data provides information of • food product characteristics, • product purchases, • household and individual characteristics. They will be used to assess the potential impacts of fiscal on added sugar and sugar content front-of-pack information policies in several European countries
Explain the relation to the objectives of the project	The home scanner data will be used to assess the potential impacts of fiscal on added sugar and sugar content front-of-pack information policies in Europe. The main output of this evaluation will be to estimate variations in the quantity of added sugar from food products purchased by, and specifically marketed to, children, and so almost exclusively consumed by children (e.g. breakfast cereals, sweetened biscuits and compotes). A specific focus will be given to socio-economically disadvantaged children. Our estimations will be based on home scanner data from several European countries (at least UK and France and preferably Portugal and Spain as well), over one year.
Specify the types and formats of data generated/collected	Csv file
Specify if existing data is being re-used (if any)	Yes
Specify the origin of the data	Unknow for the moment. Negotiation with Europanel is ongoing (November 2019).
State the expected size of the data (if known)	8 Go for French Kantar Home scanner data Unknown for the other countries
Outline the data utility: to whom will it be useful	French Kantar dataset will be used only by INRA. WP4 and WP6 partners for home scanner dataset of other selected countries

WP5

Name of the Data Set	WP5
State the purpose of the data collection/generation	To investigate behavioural interventions that might have positive impacts on children's lifestyles.
Explain the relation to the objectives of the project	WP5 has as objectives a better understanding of behavioural interventions that might improve children's decision-making regarding food and beverage choices (primarily) and activity levels (secondarily).
Specify the types and formats of data generated/collected	Survey instruments, possible grocery store scanner data, and field study behavioural measures (e.g., observations of portion sizes, food choices, etc.).
Specify if existing data is being re-used (if any)	Possibility of re-using previously collected data from cooperating organizations (e.g., VIF in France).
Specify the origin of the data	Unknown at present time.
State the expected size of the data (if known)	N/A
Outline the data utility: to whom will it be useful	Work Package 5 members, as well as the entire grant group for final report; also potential stakeholders.

Healthy Lifestyle -HLS, (WP7):

Name of the Data Set	Healthy Lifestyle (HLS)
State the purpose of the data collection/generation	The main purpose of the data is to evaluate the status and changes in general physical activity, organised sports practice, health and attitude towards participation in the Healthy Lifestyle intervention program.
Explain the relation to the objectives of the project	One of the tasks in WP7 is to evaluate of the effectiveness and efficacy of theHealthy Lifestyle intervention in Slovenia
Specify the types and formats of data generated/collected	The data is in tabular form and can be converted to appropriate format (such as .csv or similar). The database contains the data, acquired through questionnaires, filled in by children, participating in the Healthy Lifestyle intervention programme.
Specify if existing data is being re-used (if any)	The data has been used before to produce annual reports but only on the level of simple statistics (frequencies and means).
Specify the origin of the data	Slovenian Sports Office Planica in partnership with the Fitlab Institute and the Faculty of Sport, University of Ljubljana. The data gathering was administered by the authorised PE teacher on each of the participating schools.
State the expected size of the data (if known)	Around 5,000 questionnaires per year from 2011 to 2018.
Outline the data utility: to whom will it be useful	The data would be of use to policy and intervention planners.

RCT Dataset, (WP8):

Name of the Data Set	Not yet decided. Recruitment still ongoing.
State the purpose of the data collection/generation	The overall aim of this study is to assess the effectiveness, feasibility and acceptance of an overweight and obesity intervention in a socially diverse population of families. In a randomized controlled trial, we will compare two conditions for the management of pre-school overweight and obesity. Participants will be families with children aged 2-6 years (n=300) with overweight or obesity (as defined by international cut-offs, with a specific focus of recruiting children at higher risk (i.e., those from vulnerable groups such as migrants or socially disadvantaged households). The participants will be randomized to the intervention group, a parent-only group treatment program focusing on evidence-based parenting practices followed by a mHealth intervention, an app offering support to a healthy lifestyle, or to the control group i.e., standard care for overweight and obesity according to the.
Explain the relation to the objectives of the project	The STOP project will examine limitations shown by approaches to addressing childhood obesity in healthcare settings in large scale experimental studies through conducting systematic analyses, and will build on the elements that have shown the greatest promise, such as the importance of parental involvement, consistent with the role played by parental lifestyles, and the value of selecting children at higher risk (especially from vulnerable populations such as migrants and socioeconomically disadvantaged households) in a new multiple site trial.
Specify the types and formats of data generated/collected	Children: Body mass index (weight and height), waist circumference, eating behaviour, physical activity, food intake, and epigenetic and metabolic markers (blood samples). Parents: Sociodemographic data and family situation (Parents are asked to fill out a background questionnaire for the child and themselves regarding: age, sex, weight and height, health status, country of birth, employment, social network, housing, and responsibility of house chores, perceived quality of life and marital status. Questions regarding the child include: country of birth, health status, living situation, number of siblings, hours per week in preschool and sedentary behaviour) and parenting behaviour. Feasibility, attrition and acceptability of the intervention: In semi-structured interviews we will thoroughly assess facilitators and barriers of recruitment as well as attrition to the intervention and feasibility and acceptability of the care offered. Both parents and health care professionals will be interviewed by trained staff in the research group post intervention.
Specify if existing data is being re-used (if any)	No existing data is being re-used.
Specify the origin of the data	Stockholm, Sweden; Mallorca, Spain; and Timisoara, Romania
State the expected size of the data (if known)	n = 300; 100 participants will be recruited from each of the three sites.
Outline the data utility: to whom will it be useful	Researchers and healthcare professionals working with treating overweight and obesity in pre-school children. Furthermore, this data could be useful to researchers developing family-based programs targeting other non-communicable paediatric diseases (e.g., diabetes and asthma).

2. FAIR data

2.1 Making data findable, including provisions for metadata:

- **Outline the discoverability of data (metadata provision)**
- **Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?**
- **Outline naming conventions used**
- **Outline the approach towards search keyword**
- **Outline the approach for clear versioning**
- **Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how**

FAIR data:

Name of the Data Set	NCD-RisC & SES analysis- WP2
Action: 1 Making data findable, including provisions for metadata	
Outline the discoverability of data (metadata provision)	Estimates of measures of obesity and more broadly anthropometry, nationally as well as by rural and urban place of residence and SES. These will all be available for visualisation and download to anyone, in STOP or outside, from http://www.ncdrisc.org/ with no restrictions and as soon as published. So a simple open process. Data can be downloaded by country or for entire regions. And file formats are Excel/CSV and there are sufficient meta-data to identify relevant variables in the files; obviously all that is needed to access the data is a web browser.
Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?	Only aggregated data is available and is therefore not identifiable.
Outline naming conventions used	See above.
Outline the approach towards search keyword	See above.
Outline the approach for clear versioning	Only the latest versions of the data are available on the website.
Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how	N/A.

Name of the Data Set	WP3 data set (s)
Action: 1 Making data findable, including provisions for metadata	
Outline the discoverability of data (metadata provision)	Existing data is available and searchable through cohort websites, and reference made to data availability in scientific publications New data and derived variables will be returned to cohort database and accessed through their specific procedures
Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?	New data will be fully linkable to cohort databases through use of project specific IDs.
Outline naming conventions used	Variable names will be chosen according to conventions in the field. They may be renamed according to systems used within cohorts.
Outline the approach towards search keyword	The approach will be cohort specific
Outline the approach for clear versioning	Over the course of the project revisions may be necessary to GIS data. Each revision will have a version number and accompanying protocol and codebook. Only final data versions will be returned to cohorts.
Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how	Full codebooks and protocols used for data generation will be provided for all new and derived data generated

Name of the Data Set	WP8 RCT (recruitment ongoing)
Action: 1 Making data findable, including provisions for metadata	
Outline the discoverability of data (metadata provision)	Not applicable
Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?	Not applicable
Outline naming conventions used	Not applicable
Outline the approach towards search keyword	Not applicable
Outline the approach for clear versioning	Not applicable
Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how	Not applicable

2.2 Making data openly accessible:

- **Specify which data will be made openly available? If some data is kept closed provide rationale for doing so**
- **Specify how the data will be made available**
- **Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?**
- **Specify where the data and associated metadata, documentation and code are deposited**
- **Specify how access will be provided in case there are any restrictions**

Action: 2 Making data openly accessible	NCD-RisC & SES analysis- WP2
Specify which data will be made openly available? If some data is kept closed provide rationale for doing so	Data will all be available for visualisation and download to anyone, in STOP or outside, fromhttp://www.ncdrisc.org/ with no restrictions and as soon as published.
Specify how the data will be made available	Available from http://www.ncdrisc.org/ with no restrictions.
Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?	All that is needed to access the data is a web browser.
Specify where the data and associated metadata, documentation and code are deposited	These will all be available for visualisation and download to anyone, in STOP or outside, fromhttp://www.ncdrisc.org/ with no restrictions and as soon as published. So a simple open process. Data can be downloaded by country or for entire regions. And file formats are Excel/CSV and there are sufficient meta-data to identify relevant variables in the files; obviously all that is needed to access the data is a web browser.
Specify how access will be provided in case there are any restrictions	No restrictions.

Action: 2 Making data openly accessible	WP3 data set (s)
Specify which data will be made openly available? If some data is kept closed provide rationale for doing so	All new data will follow a 'guarded open access' approach. Due to consents and the personal nature of the data collected in the cohorts, it not possible to provide complete open access. However, all cohorts encourage data use by external collaborators and have their own data access procedures
Specify how the data will be made available	Data will be provided upon application to cohorts. Applications are typically reviewed by a cohort steering board who judge applications based on criteria such as use of data, data security and scientific merit.
Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?	No specific tools. Data will be provided in formats generally useable by most statistical software example .csv files.
Specify where the data and associated metadata, documentation and code are deposited	Data and metadata are deposited in cohort databases.
Specify how access will be provided in case there are any restrictions	Data will be provided upon application to cohorts

Action: 2 Making data openly accessible	TBC, WP5
Specify which data will be made openly available? If some data is kept closed provide rationale for doing so	All data will be made openly available.
Specify how the data will be made available	Uploading of files.
Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?	Unknown at present time, although SPSS is most likely software to be used.
Specify where the data and associated metadata, documentation and code are deposited	N/A
Specify how access will be provided in case there are any restrictions	N/A

Action: 2 Making data openly accessible	WP8 RCT (recruitment onoging)
Specify which data will be made openly available? If some data is kept closed provide rationale for doing so	Not applicable
Specify how the data will be made available	Not applicable
Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?	As the data for WP8 has not been collected yet, these decisions have not been made yet.
Specify where the data and associated metadata, documentation and code are deposited	As the data for WP8 has not been collected yet, these decisions have not been made yet.
Specify how access will be provided in case there are any restrictions	Not applicable

2.3 Making data interoperable:

- **Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.**
- **Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?**

Action: 3 Making data interoperable	NCD-RisC & SES analysis- WP2
Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.	The NCD-RisC website provides unrestricted access to the data set as well as links to all related publications.
Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?	Standard vocabulary is used.

Action: 3 Making data interoperable	WP3 data set (s)
Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.	The data will be highly interoperable due to the standard formats, and clear codebooks provided.
Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?	Vocabulary will be standard for the field of population research

Action: 3 Making data interoperable	TBC, WP5
Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.	Unknown at present time.
Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?	Standard vocabulary will be used.

Action: 3 Making data interoperable	WP8 RCT (recruitment ongoing)
Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.	Not applicable
Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?	Not applicable

2.4 Increase data re-use (through clarifying licenses):

- **Specify how the data will be licenced to permit the widest reuse possible**
- **Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed**
- **Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why**
- **Describe data quality assurance processes**
- **Specify the length of time for which the data will remain re-usable**

Action: 4 Increase data re-use (through clarifying licenses):	NCD-RisC & SES analysis- WP2
Specify how the data will be licenced to permit the widest reuse possible	The NCD-RisC website provides unrestricted access to the data as soon as it is published.
Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed	Data will be made available for re-use immediately upon publications of results, listed as project deliverables.
Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why	Yes.
Describe data quality assurance processes	All data sets are published alongside peer-reviewed publications.
Specify the length of time for which the data will remain re-usable	Data will remain available into the foreseeable future

Action: 4 Increase data re-use (through clarifying licenses):	WP3 data set (s)
Specify how the data will be licenced to permit the widest reuse possible	The data will not be licensed
Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed	Data will be made available for re-use immediately upon publications of results, listed as project deliverables.
Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why	Data will be made available for re-use for academic purposes.
Describe data quality assurance processes	Gut hormone measurement quality will be controlled through well-established procedures including running quality control samples throughout each analytical run to assess technical variation and analytical drift and if necessary correct or exclude low quality measurements. Metafiles will provide data quality statistics for each measurements (e.g. coefficients of variation) and flags for low quality data. Detailed data quality checks following established standard procedures will be performed both on existing data sources and newly derived data. Detailed checks for inconsistencies and repetitions will have been performed by the data manager for the specific cohort studies and no issues are expected.
Specify the length of time for which the data will remain re-usable	Cohorts have independent funding for maintaining their databases into the foreseeable future

Action: 4 Increase data re-use (through clarifying licenses):	TBC, WP5
Specify how the data will be licenced to permit the widest reuse possible	Unknown at present time.
Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed	No data embargo will be needed.
Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why	Data can be used by third parties after the end of the project.
Describe data quality assurance processes	Unknown at present time
Specify the length of time for which the data will remain re-usable	Indefinitely

Action: 4 Increase data re-use (through clarifying licenses):	WP8 RCT (recruitment ongoing)
Specify how the data will be licenced to permit the widest reuse possible	Not applicable
Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed	Not applicable
Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why	Not applicable
Describe data quality assurance processes	Not applicable
Specify the length of time for which the data will remain re-usable	The data collected in this randomized control trial can be used for meta-analyses in the future.

3. Allocation of resources

Explain the allocation of resources, addressing the following issues:

- **Estimate the costs for making your data FAIR. Describe how you intend to cover these costs**
- **Clearly identify responsibilities for data management in your project**
- **Describe costs and potential value of long term preservation**

Name of the Data Set	**NCD-RisC & SES analysis- WP2**
Action: Explain the allocation of resources, addressing the following issues	
Estimate the costs for making your data FAIR. Describe how you intend to cover these costs	Availability of NCD-RisC data pre-dates the STOP project.
Clearly identify responsibilities for data management in your project	The WP leader, Majid Ezzati, will have ultimate responsibility for data management
Describe costs and potential value of long term preservation	Not applicable.

Name of the Data Set	WP3 data set (s)
Action: Explain the allocation of resources, addressing the following issues	
Estimate the costs for making your data FAIR. Describe how you intend to cover these costs	Costs will be met by cohort funds
Clearly identify responsibilities for data management in your project	Task leaders responsible for production of new data will have responsibility for providing clear codebooks and quality assurance of data. The WP leader (Paolo Vineis) will have ultimate responsibility for data management
Describe costs and potential value of long term preservation	Not applicable as data will be stored at cohort studies.

Name of the Data Set	TBC, WP5
Action: Explain the allocation of resources, addressing the following issues	
Estimate the costs for making your data FAIR. Describe how you intend to cover these costs	Unknown at present time, but STOP funds will be used, if necessary.
Clearly identify responsibilities for data management in your project	WP5 leader: Tina M. Lowrey
Describe costs and potential value of long term preservation	Unknown at present time.

Name of the Data Set	WP8 RCT (recuitment ongoing)
Action: Explain the allocation of resources, addressing the following issues	Not yet decided.
Estimate the costs for making your data FAIR. Describe how you intend to cover these costs	Not applicable.
Clearly identify responsibilities for data management in your project	Paulina Nowicka will be responsible for overall data management in WP8. As data collection has not begun yet this cannot be more detailed.
Describe costs and potential value of long term preservation	Not applicable.

4. Data security

Address data recovery as well as secure storage and transfer of sensitive data

Name of the Data Set	NCD-RisC & SES analysis- WP2
Action: Address data recovery as well as secure storage and transfer of sensitive data	Data will be stored on the servers at Imperial College which are backed-up weekly in an ISO 27001 secure environment at Imperial College.

Name of the Data Set	WP3 data set (s)
Action: Address data recovery as well as secure storage and transfer of sensitive data	Data will be stored on the servers at Imperial College which are backed-up weekly in an ISO 27001 secure environment at Imperial College. Data will be transferred through sFTP such as the Imperial College file exchange service.

Name of the Data Set	TBC, WP5
Action: Address data recovery as well as secure storage and transfer of sensitive data	N/A

Name of the Data Set	WP8 RCT (recruitment ongoing)
Action: Address data recovery as well as secure storage and transfer of sensitive data	The transfer of data between the three participating universities in this project will be done after data sharing agreements are signed. All data sharing agreements will follow the rules and regulations of the involved institutions.

5. Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

Name of the Data Set	NCD-RisC & SES analysis- WP2
Action: To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former	Data is not identifiable and only provided in aggregate form.

Name of the Data Set	WP3 data set (s)
Action: To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former	The main risk is that identifiable data on participants may spill into an uncontrolled environment. This will be minimised by only using pseudonymised data (with identifiable information held separately within the cohorts) and by developing an information governance policy which will be reviewed by the School of Public Health data officer to ensure compliance with GDPR. All data will be strictly access controlled to named researchers only, who have received data security training.

Name of the Data Set	TBC, WP5
Action: To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former	N/A

Name of the Data Set	WP8 RCT (recuitment ongoing)
Action: To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former	Each participating university (Karolinska Institutet, Huddinge, the University of the Balearic Islands & CIBEROBN, and the University of Medicine and Pharmacy "Victor Babes") has received ethical approval for this multi-site randomized controlled trial. The treatment of personal data is in accordance with the EU General Data Protection Regulation (GDPR), which means that unauthorized people cannot access the information. Data will be saved for 10 years after the completion of the study and thereafter it will be deleted. Karolinska Institutet, Huddinge, the University of the Balearic Islands & CIBEROBN, and the University of Medicine and Pharmacy "Victor Babes" is responsible for managing the data and privacy of personal information for the Swedish, Spanish, and Romanian data, respectively. According to GDPR, participants have the right to once every calendar year (without any cost), receive information about their personal information and how we handle this information. If necessary, participants can get any errors corrected. Participants can also request deletion of personal information and that treatment of their information is limited. If you participants choose to withdraw from the study, no more information will be collected, but we are entitled to retain the information already collected. In such cases, these data will be encoded so that they can in no way be linked back to the participant.

6. Other

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

Name of the Data Set	**NCD-RisC & SES analysis- WP2**
Action: Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)	Data Management Policy & Procedures https://www.imperial.ac.uk/media/imperial-college/research-and-innovation/research-office/public/Imperial-College-RDM-Policy.pdf Data Security Policy http://www.imperial.ac.uk/admin-services/secretariat/college-governance/charters/policies-regulations-and-codes-of-practice/information-security-/policy/

Name of the Data Set	WP3 data set (s)
Action: Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)	Data Management Policy & Procedures https://www.imperial.ac.uk/media/imperial-college/research-and-innovation/research-office/public/Imperial-College-RDM-Policy.pdf Data Security Policy http://www.imperial.ac.uk/admin-services/secretariat/college-governance/charters/policies-regulations-and-codes-of-practice/information-security-/policy/ Data Sharing Policy Institutional: http://www.environment-health.ac.uk/information-governance-and-data-management-policy HELIX cohort: https://www.projecthelix.eu/files/helix_external_data_request_procedures_final.pdf Institutional Information Policy http://www.imperial.ac.uk/admin-services/secretariat/college-governance/charters/policies-regulations-and-codes-of-practice/policy-framework/

Name of the Data Set	TBC, WP5
Action: Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)	N/A

Name of the Data Set	WP8 RCT (recruitment ongoing)
Action: Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)	Recruitment is still ongoing. Details confirmed at a later date.