

Pro**PASS**

Prospective Physical Activity, Sitting, and Sleep Consortium

**Copenhagen Workshops**

Thursday 11th – Friday 12th October 2018

The National Research Centre for the Working Environment

Lersø Parkallé 105, 2100 København, Denmark

# **Table of Contents**

[Table of Contents 2](#_Toc526878758)

[1. Program 3](#_Toc526878759)

[2. Information Brochure 5](#_Toc526878760)

[3. List of Delegates 8](#_Toc526878761)

[4. Report of the Advisory Committee 9](#_Toc526878762)

[5. Aims of the Workshops 10](#_Toc526878763)

[6. Workshop 1 11](#_Toc526878764)

[7. Workshop 2 15](#_Toc526878765)

[8. Workshop 3 17](#_Toc526878766)

[9. Workshop 4 19](#_Toc526878767)

[10. Workshop 5 21](#_Toc526878768)

[11. Workshop 6 24](#_Toc526878769)

[References 28](#_Toc526878770)

# **Information Brochure**

**What is Pro PASS?**

An international consortium of thigh-worn accelerometry data that quantify physical activity and posture. ProPASS is a platform for pooled individual participant data analyses and prospective meta-analyses of epidemiological studies that examine the health effects of detailed physical activity, sleep, and postural patterns.

**About the Project**

ProPASS is a consortium of epidemiological studies that *have collected* or *plan to collect* data using 3-axial sensors that are attached to the thigh. Such methods are feasible and are increasingly used in health research thanks to their capacity to record information on physical activity volumes as well as posture allocation data simultaneously. Minimum study entry requirement is the use of *one thigh worn (3-axial) sensor* with the capacity to output raw acceleration.

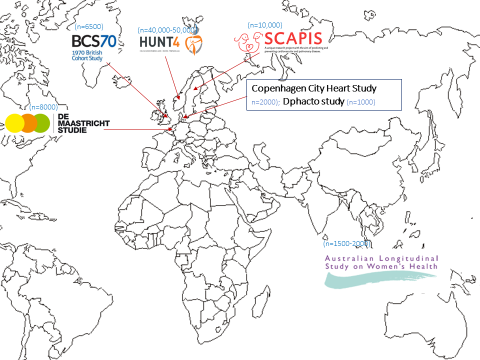
The vision for ProPASS is to be an authoritative shared data resource for expanding scientific understanding of how physical activity and posture affect health outcomes.

The ProPASS objectives are:

1. To develop an international network of consortium partners to contribute data and/or resources for the advancement of the consortium;
2. To develop a set of tools for pooling, harmonising, and analyzing data of existing studiestowards individual participant meta-analytical projects;
3. To develop a platform for prospective meta-analysis through developing a set of tools for data collection protocols to be used by future studies in the field;
4. To be a dynamic and steadily growing resource that will be able to answer fine-resolution research questions in ways not previously possible.

ProPASS is developed in a way that can accommodate further advancements, such as addition of data from other placement sensors.

**Participating cohorts** (as of June 2018)



**ProPASS Working Group**

|  |  |  |  |
| --- | --- | --- | --- |
|  | PROF EMMANUEL STAMATAKIS  Charles Perkins Centre  The University of Sydney |  | PROF ANDREAS HOLTERMANN  National Centre For The Working Environment, Copenhagen |
|  | PROF MARK HAMER  National Centre for Sport & Exercise Medicine  Loughborough University |  | ASSOCIATE PROF VEGAR RANGUL  The Norwegian University Of Science & Technology |

**ProPASS Advisory Group**

|  |  |  |  |
| --- | --- | --- | --- |
|  | DR CHARLES MATTHEWS  National Cancer Institute,  Cancer Epidemiology & Genetics  National Institute of Health | C:\Users\tcoo3501\Desktop\Lee headshot.jpg | PROF I-MIN LEE  Division of Preventive Medicine,  Department of Medicine  Harvard University |
|  | PROF LISA ASKIE  NHMRC Clinical Trials Centre  Cochrane,  The University of Sydney |  | PROF ULF EKELUND  Norwegian Institute of Public Health  Norwegian School of Sport Sciences |
|  | PROF PETER CISTULLI  Charles Perkins Centre &  Northern Clinical School  The University of Sydney |  | PROF ADRIAN BAUMAN  Prevention Research Collaboration  Charles Perkins Centre & School of Public Health  The University of Sydney |

**ProPASS Current Collaborators**

|  |  |  |  |
| --- | --- | --- | --- |
|  | PROF METTE AADAHL  Centre for Clinical Research and Prevention, Denmark |  | PROF ANNEMARIE KOSTER  School of Public Health & Primary Care  Maastricht University |
|  | PROF ODDGEIR L HOLMEN  The Nord-Trondelag Health Study  The Norwegian University of Science & Technology |  | PROF PAUL JARLE MORK  Department of Public Health & Nursing  The Norwegian University of Science & Technology |
|  | DR PETER PALM  SCAPIS Project  Uppsala University |  | DR SEBASTIEN CHASTIN  Glasgow Caledonian University |
|  | PROF MALCOLM GRANAT  School of Health Sciences  University of Salford |  | DR ANNA PULAKKA  Institute of Clinical Medicine  University of Turku |

**Commercial Partners**

|  |  |
| --- | --- |
|  | PAL Technologies  Glasgow, UK |

# **List of Delegates**

|  |  |  |
| --- | --- | --- |
| **Name** | **Institution** | **Email** |

**Attendees**

|  |  |  |
| --- | --- | --- |
| Adrian Bauman | University of Sydney | [Adrian.bauman@sydney.edu.au](mailto:Adrian.bauman@sydney.edu.au) |
| Andreas Holtermann | National Research Centre for the Working Environment, Copenhagen, University of Southern Denmark | [aho@nfa.dk](mailto:aho@nfa.dk) |
| Anna pulakka | University of Turku, Finland | [anna.pulakka@utu.fi](mailto:anna.pulakka@utu.fi) |
| Annemarie Koster | Maastricht University | [a.koster@maastrichtuniversity.nl](mailto:a.koster@maastrichtuniversity.nl) |
| Cecilie Goltermann Funder | Centre for Clinical Research and Prevention, Denmark | [Cecilie.Goltermann.Funder@regionh.dk](mailto:Cecilie.Goltermann.Funder@regionh.dk) |
| Douglas Maxwell | PAL Technologies | [douglas@paltechnologies.com](mailto:douglas@paltechnologies.com) |
| Lisa Askie | Cochrane & University of Sydney | [lisa.askie@ctc.usyd.edu.au](mailto:lisa.askie@ctc.usyd.edu.au) |
| Malcom Granat | International Society for the Measurement of Physical Behaviour, University of Salford, PAL Technologies | [malcolm@paltechnologies.com](mailto:malcolm@paltechnologies.com) |
| Manos Stamatakis | University of Sydney | [emmanuel.stamatakis@sydney.edu.au](mailto:emmanuel.stamatakis@sydney.edu.au) |
| Mark Hamer | Loughborough University | [m.hamer@lboro.ac.uk](mailto:m.hamer@lboro.ac.uk) |
| Matthew Stevens | National Research Centre for the Working Environment, Copenhagen | [mws@nfa.dk](mailto:mws@nfa.dk) |
| Mette Aadahl | Centre for Clinical Research and Prevention, Denmark | [mette.aadahl@regionh.dk](mailto:mette.aadahl@regionh.dk) |
| Patrick Crowley | National Research Centre for the Working Environment, Copenhagen | [pjc@nfa.dk](mailto:pjc@nfa.dk) |
| Paul Jarle Mork | Norwegian University of Science & Technology | [paul.mork@ntnu.no](mailto:paul.mork@ntnu.no) |
| Peter Palm | Uppsala University Hospital | [peter.palm@akademiska.se](mailto:peter.palm@akademiska.se) |
| Sari Stenholm | University of Turku, Finland | [samast@utu.fi](mailto:samast@utu.fi) |
| Vegar Rangul | Norwegian University of Science & Technology | [vegar.rangul@ntnu.no](mailto:vegar.rangul@ntnu.no) |

# **Aims of the Workshops**

**Aims & Deliverables**

* Meet and become acquainted with members of the consortium
* Agree on a core set of variables for harmonisation and outcomes of the consortium
* Develop a protocol for ProPASS
* Develop clear procedures for data access, sharing, processing, and linkage
* Publish workshop/conference proceedings
* Register protocol on PROSPERO
* Agree on firm timelines for ProPASS
* Plan next steps for action

# **Workshop 1**

Selection for Harmonisation of Non-Accelerometry Variables

1. **Introduction**

The umbrella scientific aim of the ProPASS consortium is to examine the independent and joint associations of the physical activity-related components of the 24 hour continuum (physical activity, sleep, time patterns, volume, frequency, intensity) with general and specific health outcomes. Analyses addressing such research questions will often involve variables on sociodemographic descriptors, general and specific health status and disability, confounders, effect modifiers (moderators), mediators, and health outcomes. Data harmonisation involves achieving or improving comparability of similar measures collected by different studies. The key to the harmonisation process is the judgement on whether data from contributing studies are ‘inferentially equivalent’ (Atkin, et al., 2017), meaning that the constructs assessed are sufficiently comparable in format, structure, or meaning. This requires consideration of whether data should be combined *and* whether data can be combined.

The aims of this session are:

1. To compile a preliminary list of non-accelerometry variables to be harmonised across ProPASS cohorts;
2. To provide an initial assessment of the harmonisation potential of these variables, based on other physical activity consortia efforts and the collective experience of ProPASS collaborators.
3. **Harmonisation In Other Physical Activity Consortia**

Here we provide two examples from the field of physical activity:

1. The European Determinants of Diet and Physical Activity Knowledge Hub (**DEDIPAC**);
2. International Children’s Accelerometry Database (**ICAD**).

DEDIPAC (Lakerveld, et al., 2017) was a project designed to enable comprehensive and large-scale research using the wide variation in exposures and outcomes that exist across Europe. It focused on correlates and determinants of physical activity and sedentary behaviour, and was developed as a 5-step process of development: (1) identification of relevant datasets across Europe; (2) development of a compendium including details on the design, study population, measures and level of accessibility of data from each study; (3) definition of key topics and approaches for secondary analyses; (4) process of gaining access to datasets; and (5) the development of a data harmonisation platform for data pooling. Out of 104 identified data sources with self-reported or objectively measured physical activity and its correlates, 14 studies were pooled for the purposes of a simulation exercise (Lakerveld, et al., 2017). The authors concluded that harmonisation was often restricted to just a few core (crude) outcome variables and some individual-level sociodemographic correlates of these behaviours.

ICAD is an accelerometry consortium of children and adolescents wearing waist-mounted Actigraphs. The selection of variables to be harmonised in ICAD was highly systematic: First, based on specific research questions of interest to the Working Group and constructs which were considered to be most valuable to a wide range of prospective users of the updated ICAD database (e.g. as confounding variables). The harmonisation of the non-accelerometry variables was decentralised, i.e. it was done by the ICAD collaborators/data contributors using ICAD developed protocols.

The first iteration of ICAD took place in 2008-2010. To date, the second iteration of ICAD iteration (ICAD 2) has pooled data from over 37,000 participants from Europe, North and South America and Australia. In its first few years, ICAD published pooled analyses involving relatively few outcome variables (waist circumference, triglycerides, glucose, blood pressure, and HDL cholesterol (Ekelund, et al., 2012)). ICAD 2 has expanded its list of variables that are harmonised or are currently being harmonised by the ICAD collaborators/data contributors (see **Appendix 1.1**). The updated list includes:

* Height, weight (harmonised across *all* collaborating studies)
* Waist circumference, skinfolds (harmonised in *about 1/2* of collaborating studies)
* Cardiometabolic health outcomes (blood pressure, glucose, insulin, HDL, cholesterol, triglycerides, etc.) (harmonised in *about 1/3 to 1/2* of collaborating studies)
* Parental education (harmonised in about 1/2 of collaborating studies)
* Parental employment (harmonised in *about 1/2* of collaborating studies)
* Ethnicity parental education (harmonised in *about 1/2* of collaborating studies)
* Car ownership (harmonised in *about 1/3* of collaborating studies)
* Sleep (harmonised across *in 1/2* of collaborating studies)
* Questionnaire-based PA, SB, transportation to and screen time (harmonised in *about 1/4 to 1/2* of collaborating studies)
* Dietary indicators (harmonised in *about 1/2* of collaborating studies)

It is worth noting that the large majority of harmonised/currently being harmonised ICAD 2 variables are available in *half or fewer contributing studies*, which indicates a focus on inclusiveness rather than completeness of the harmonised dataset. ICAD has published a detailed paper describing the methods, processes, and challenges involved when harmonising physical activity correlates data (Atkin, et al., 2017). An ICAD 2 general principle was the creation of multiple harmonised variables for each construct, balancing the often competing demands of resolution and coverage (number of included studies). This enabled the creation of higher resolution variables that made best use of detailed data where it was available and lower resolution variables that allowed for inclusion of the largest number of studies in the analyses.

1. **Non-Physical Activity Harmonisation Projects: CLOSER – Cohort and Longitudinal Studies Enhancement Resources**

CLOSER (UCL, 2018) is a project aimed at bringing together eight major UK longitudinal studies. It is managed by the Institute of Education at University College London. Data harmonisation is one of the major aspects of CLOSER’s activity and it has been divided into a number of projects, with each one concerning a family of variables. The following CLOSER harmonisation sub-projects are the ones that are most relevant to ProPASS:

* Harmonising measures of body size and body composition;
* Harmonising measures of occupation and education;
* Harmonising earnings and income;
* Harmonising strategies for analysing biological samples;
* Exploiting the existing biomarker data available in CLOSER;
* Socioeconomic differentials in physical activity by age and cohort.

For each of these harmonisation projects, there are detailed project descriptions in the CLOSER website, as well as user guides, and workshop summary proceedings. CLOSER could assist our harmonisation efforts in a number of ways, for example:

* Use CLOSER as a workflow and methodology model;
* Use some of the resources that CLOSER has put in the public domain;
* Establish direct collaborative links and, if possible, involve CLOSER some aspects of the ProPASS harmonisation projects.

1. **Literature Review Of Adult Thigh-Worn Accelerometry**

We have carried out a systematic set of searches of studies that used thigh-worn accelerometers in adults. Following study eligibility assessment (see **Appendix 1.2** for protocol overview) 40 studies were selected (see **Appendix 1.3** for full list of selected studies). For the purposes of this Workshop session, we have extracted data from 7 randomly selected studies:

* Bellettiere, J, et al (2017). Associations of sitting accumulation patterns with cardio-metabolic risk biomarkers in Australian adults. *PLoS ONE*, 12(6): e0180119. doi: 10.1371/journal.pone.0180119.
* de Rooij, BH, et al. (2016). Physical Activity and Sedentary Behaviours in Metabolically Healthy versus Unhealthy Obese and Non-Obese Individuals - The Maastricht Study. *PLoS ONE*, 11(5): e0154358.
* Denkinger, MD, et al. (2010). Accelerometer-based physical activity in a large observational cohort-study protocol and design of the activity and function of the elderly in Ulm (ActiFE Ulm) study. *BMC Geriatrics*, 10: 50-64.
* Gupta, N, et al. (2015). Is Objectively Measured Sitting Time Associated with Low Back Pain? A Cross-Sectional Investigation in the NOMAD study. *PLoS ONE*, 10(3): e0121159. doi:10.1371/journal.pone.0121159.
* Lagersted-Olsen, J, et al. (2016). Does objectively measured daily duration of forward bending predict development and aggravation of low-back pain? A prospective study. *Scandinavian Journal of Work, Environment & Health*, 42(6): 528-537.
* Shaw, RJ, et al. (2017). The Influence of Neighbourhoods and the Social Environment on Sedentary Behaviour in Older Adults in Three Prospective Cohorts. *International Journal of Environmental Research & Public Health*, 14(6): 24.
* Smith, L, et al. (2013). Active buildings: modelling physical activity and movement in office buildings. An observational study protocol. *BMJ Open*, 3: e004103. doi:10.1136/bmjopen-2013-004103.

**Table 1** (**Appendix 1.4**) lists the non-accelerometry variables available in the above 7 studies of thigh-worn accelerometry. In terms of potential confounders (e.g. socioeconomic, occupational, lifestyle health behaviours) the preliminary data extraction shows that there are few overlapping constructs among studies, and these constructs are measured in ways that may not be directly harmonisable. Among biological outcomes, only BMI and waist circumference are almost universally present. Several other cardio metabolic outcomes (glucose, cholesterol, Hb1Ac, etc.) are present in about half of the studies with seemingly good harmonisation potential.

1. **Points for Discussion**

Previous physical activity consortia efforts, as well as our preliminary assessment of the thigh-worn accelerometry literature show that the selection of non-accelerometry variables to be harmonised needs to be thorough and methodical. ProPASS could potentially benefit from considering to adopt the following principles learnt from the ICAD and DEDIPAC harmonisation experience:

* Expect that the entire retrospective harmonisation of the non-accelerometery variables will require substantial resources and a well thought out plan
* Follow a systematic process for variable selection for harmonisation
* Develop an initial list of variables to harmonise *to fulfil the needs of the list of research questions ProPASS will address*
* Utilise a model of variables selection for harmonisation on the grounds of their importance, not universal availability. See for example the ‘parental TV time’ and ‘availability of TV in the bedroom’ variables were selected to be harmonised, despite their availability in a small minority of ICAD studies.
* Rigorous conduct and reporting of retrospective data harmonisation, which will enable better analysis and interpretation of derived data.
* Consider harmonisation of multiple version of the same construct (e.g. ethnicity) with each variable containing a different resolution of information (e.g. ‘white-other’ vs. ‘white-Chinese-Black’ vs. ‘African-Black’ vs. ‘Caribbean-South Asian’).
* The challenges with retrospective harmonisation noted by ICAD and the thigh-worn accelerometery literature on the whole lend support to ProPASS investing on prospective harmonisation, whenever possible (**Workshop 3**).
* Consider the CLOSER model of harmonisation and its methods
* Discuss collaboration with other projects that have first-hand expertise on cohort data harmonisation, like CLOSER and ICAD.

# **Workshop 2**

Methods of Variable Harmonisation

1. **Introduction**

Heterogeneity between studies in the design, selection criteria, data collection and measures collected limits our capacity to easily compare or integrate data.

***Data Harmonisation involves achieving or improving comparability of similar measures*** collected by separate studies or databases for different individuals. Guidelines for conducting rigorous data harmonisation are provided by the Maelstrom Research group (Fortier, et al., 2015).

**Example:**

The *international children’s accelerometry database (ICAD)* is an excellent example of a recent application of the Maelstrom guidelines. This consortium of 20 partners shared raw accelerometer data and associated non-accelerometry data. This data was processed using standardized methods facilitating the creation of harmonized and comparable variables of over 37,000 young people, between the ages of 3 and 18 years old. This data comes from studies across Europe, the US, Brazil, and Australia and is available for use, on request, within 5 to 10 working days. ICAD initially began with *retrospective harmonisation* but has recently begun to develop/encourage *prospective harmonisation*.

1. **Steps for Data Harmonisation**
2. Identify relevant studies/cohorts
   * Currently being conducted through scoping review and ProPASS member surveys.
3. Define core variables to be harmonised (DataSchema)
   * Discuss and come to decisions on **what are the important variables requiring harmonisation**. This must be **based upon what variables are available** in the relevant studies/cohorts (see **Workshop 1**).
4. Evaluate harmonisation potential
   * Evaluate the potential for harmonisation by **considering the appropriateness of each of the processing methods** provided below for harmonising each core variable, the consequences that this will have on the core variable and **decide whether this consequence is acceptable**.
5. Process data
   * Conduct harmonisation of the DataSchema using the appropriate method.
6. **Data Processing Options**

*Algorithmic transformation*

**Definition**: A finite set of unambiguous instructions performed in a prescribed sequence to achieve a goal, especially a mathematical rule or procedure used to compute a desired result.

**Application**: To harmonise same measures (continuous variables, categorical, or both) with different but combinable ranges or categories.

**Example**: Processing ‘number of cigarettes per day’ in numerical form into ‘number of cigarettes per day’ in categorical form (1: <10 cigarettes, 2: 10-20 cigarettes, 3: 21+ cigarettes).

**Advantages/Disadvantages**:The **a**dvantage of this approach is that it allows for analysis by stratification, making it easy to identify more extreme groups. A disadvantage of this however is the loss of information on an individual level.

*Simple calibration model*

**Definition**: A mathematical model that transforms one continuous measure into another continuous measure (or vice versa) to operate at the same unit.

**Application**: To harmonise same metric measures using different scales with calibration model. The calibration model can either be known or estimated from data in case multiple measure are measured on a representative set of participants.

**Example**: Processing ‘height, measured in cm into ‘height’ in inches.

**Advantages/Disadvantages**: No clear advantages/disadvantages, however this type of data is usually only anthropometrical.

*Standardization model*

**Definition**: A model that centralizes, standardizes, or normalizes data to bring all variables into proportion with one another, with or without stratification or regression of other variables.

**Application**: To harmonise same constructs measured using different scales with no known calibration method or bridging items. Normalization is typically used for discrete interval measures.

**Example**: Processing memory measures with no common items into standardized measure like C-scores or T-scores.

**Advantages/Disadvantages**: The advantage of this approach is that you get normal distributions for variables with otherwise tricky distribution; however this also has a disadvantage in that you may lose some information contained in these distribution trends.

*Latent variable model*

**Definition**: A statistical model that relates a set of actual variables to a set of latent variables. Latent variables are constructs that are not directly observed but are rather inferred (through mathematical model) from actual variables that are observed (directly measured).

**Application**: To harmonise same constructs measured using different scales with no known calibration method but with bridging items present.

**Example**: Processing memory measures with common items according to latent variable model.

**Advantages/Disadvantages**: The advantage of this model is that it makes it possible to compare items based on the underlying construct used to develop the measure, although this also requires a thorough understanding and agreement on the underlying constructs, which often rely on subjective decisions.

*Multiple imputation models*

**Definition**: A statistical technique for imputing missing values with a set of plausible values that represent the uncertainty about the right values to impute.

**Application**: To harmonise datasets (and not variables) with the exact same set of variables using bridging variables.

**Example**: Processing activities of daily living measures by imputing missing items.

**Advantages/Disadvantages**: Advantages of this method lie in the ability to use data with missing values, although perhaps more disadvantages in the reasoning behind the selection of inputted values, the acceptable threshold of missing values, and the implications for future analysis. This method must be applied with caution due to the risk of biasing.

See **Table 2** for further details.

# **Workshop 3**

Development of harmonised protocols for future physical activity studies

* 1. **Introduction**

**Prospective meta-analyses and Individual Participant Data meta-analyses**

A prospective meta-analyses (PMA) refers to the design of multiple trials with the explicit, predefined purpose to, when completed, combine them in a meta-analysis. Depending on the breadth of the questions answered, prospective, pre-planned meta-analyses can either address focused questions or shape large research agendas for a long-term research programme or even an entire discipline (Ioannidis, 2017). PMA have features in common with both cumulative meta-analyses and those involving Individual Participant Data (IPD) (Cochrane, 2018). PMA can help overcome some of the recognized problems of retrospective meta-analyses (Cochrane, 2018):

* Enabling hypotheses to be specified a priori ignorant of the results of individual studies;
* Enabling prospective application of study selection criteria;
* Enabling a priori statement of intended analyses to be made before the results of the individual studies are known.

Hence, key benefits of prospective meta-analyses are that pre-planned analyses will be performed with widely scrutinised, pre-specified high-quality standards, and there will be no selective reporting of the results. PMA are very uncommon to date, possibly due to the extensive co-ordination required (Ioannidis, 2017).

In observational studies IPD allows for adjustment of confounding factors in the meta-analyses and the main advantage of IPD meta-analysis is that researchers can assess the influence of participant-level covariates on all collected outcomes and measured time points of interest.

**Prospective harmonization**

Integration or comparison of individual participant data requires the generation of compatible, i.e. harmonized datasets across studies. For ***prospective harmonization***, study investigators will agree upon common measures and protocols before beginning data collection. Agreement on a core set of common measures and collection procedures ***prior to data collection*** will facilitate future integration or comparison of data across standalone studies. Within the field of physical activity epidemiology, no prospective harmonization of data has yet been undertaken, but within other fields of research, e.g. measurement of cognitive functioning, researchers have started the process of harmonizing data prospectively, while also developing statistical methods for harmonization of retrospective data (Griffith, et al., 2016) (Griffith, et al., 2013).

Whereas prospective harmonization of accelerometer data, i.e. dissemination of a standardized protocol for collection of accelerometer data, may be possible; prospective harmonization of contextual data and health outcomes may be scientifically and technically challenging. See **Appendix 3.1** for an example of a harmonised protocol of accelerometry data.

* 1. **Suggestions for contextual and health outcome variables for harmonization in future studies**

To facilitate future pooling of data for prospective meta-analyses, harmonization of methods for data collection of contextual and health outcome variables is warranted (Atkin, et al., 2017) (Fortier, et al., 2015).

See **Table 3** for core contextual variables, and **Table 4** for core intermediate health outcomes (**Appendix 3.2**)

* 1. **Options for discussion**
* Challenges with convincing future studies to use a set methodology to measure accelerometry?
* Challenges with convincing future studies to use a set methodology to measure health outcomes and covariates?
* What kind of studies do we expect to be able to harmonize prospectively
  + Smaller local cohorts?
  + New cross-sectional studies?
  + Is it realistic to expect to be able to influence large cohorts?
* Overall: is it going to be possible at all and worth the effort?
* Strategy to identify and engage with future studies as a consortium?

# **Workshop 4**

ProPASS Accelerometry Data Processing Methods

1. **Data Processing Options**
2. Commercial software:
   * Generally provide the options of exporting a unique file format, a general .csv format or an ‘event’ file (divided into epochs).
   * ActiLife and variations of the ActivPAL software are the most frequently cited commercial software in use, in this review.
   * PAL analysis software provides inclination-based information about sitting, standing, walking, and sitting bout duration, and accelerometer orientation and volume.
   * ActiLife analysis is based primarily on counts-per-minute, providing information about metabolic equivalents, energy expenditure, and sleep behaviours.
3. Rule based software:
   * The ‘rule-based’ in this case refers to the classification of activities by developing rules which the data must fulfil in order to be classified as a given activity (**Figure 1, Appendix 4.1**).
   * Acti4 was the most frequently cited rule-based software in use, in this review.
   * Acti4 uses this rule-based approach (**Figure 1**) to classify a number of daily activities and postures, including; sitting, lying, moving, walking, walking on stairs, cycling, and running
4. Machine-learning approach:
   * Classifies physical activity based on activity specific features identified in the data.
   * These features can be classified into groups and applied to other datasets to automatically classify activities. The ‘learning’ aspect stems from the idea that the classification of activities becomes more and more precise through feedback loops.
   * One such example is a Random Forest Classifier (RFC). Overall the RFC method is a two stage approach:
     + 1) identifying a number of activity specific features at random in the dataset and iterating this process until there are a sufficient number of identified features for classification
     + 2) using these features to predict outcomes in other datasets.

See **Table 5**, **Table 6**, and **Table 7** for the pros and cons of data processing options (**Appendix 4.2**).

1. **Comments for discussion:**

Although often unreported in the reviewed studies, commercial software is almost always necessary to initialize and download data. They also provide an easy analysis tools and an often user-friendly analysis option. Moreover, the cost and time commitment to developing and maintaining the software is the burden of the company and not the researcher. However, there is little control over how the data is treated, what outcome measures are retrieved, and what algorithm/cut-off point/classification technique is implemented. These aspects are still the topics of much debate in the research community – cut off points being a prime example - and fundamental aspects to control for analysis. Therefore, unless there is a more transparent process in place where proprietary analysis methods are detailed precisely, it is difficult to promote the use of current commercial software for ProPASS purposes.

Acti 4 is software developed at the National Research for the Work Environment, Copenhagen (NFA) and used in many subsequent studies conducted by researchers at the NFA and some collaborators. It has the advantages of an open-source license, a transparent analysis process, and activity classifications that are based on postures and movement as opposed to the more problematic counts-per-minute. For these reasons, Acti4 would seem a good alternative to commercial software packages. However, the current version requires some training before it can be used, mainly to learn how to correctly synchronize the recordings from multiple sensors and create time intervals based on the diary information (e.g. 8:00 to15:30 was work time, 15:30 to 22:30 was leisure time etc.). This process takes approximately 15-20mins for a given subject with one full week of accelerometer recording and diary information. This estimate can vary considerably depending on; the quality of recording and the detail/length of the diary. The synchronization in this case refers to ensuring that sensors placed on the thigh and upper back, or other locations, are in synch and showing the same events/incidents in time. The inclination of the accelerometers must also be verified throughout the recording, to ensure they remain placed within a reference orientation. It must also be considered that since research questions often differ slightly, the software must often be customised for this purpose, which also takes time.

The review provided less information about the machine-learning approach. The great potential of this approach lies in bypassing the more time consuming processes of the Acti4 program. In theory, it would be a fully automated process, with high accuracy and efficient for larger data sets. It also requires considerable resources and specific knowledge of machine learning algorithms and human movement/postures to develop it to the point where it can serve as a shared platform for analysis of data from various accelerometer manufacturers.

In summary, Acti4 appears to be the most promising approach in the short- to medium-term, provided that it can be modified to follow a less time consuming process to not rely on the learning procedures outlined above. This could be achieved through automating the detection of synchronization events and digitalizing diary entries, for example. Over the longer-term, and for use on very large cohorts, the machine learning approach is considered the best suited approach (for ProPASS). Also important to consider are combinations of the two analytical approaches. For example, if the two approaches provide very similar physical activity classifications, datasets using either Acti4 or machine learning could be harmonized. In addition to the pros/cons highlighted above, it is important to note that the resource demands, both in terms of time and money, for both approaches are difficult to estimate but can be assumed to be considerable.

**Table 8** summarised the data processing methods used in the review studies (**Appendix 4.3**).

# **Workshop 5**

Linkage of ProPASS data with health outcomes

(Public Health Research Data Forum, March 2015)

1. **Key summary points**

* Linked mortality data are available in most ProPASS studies; hospitalisation data have good potential.
* We know little about how the quality/validity of data linkage varies between studies. We need to better understand some of the key institutional issues of individual studies.
* Pooling linked data sets across countries appears feasible in the majority of cases, although substantial variation may exist in the approach to obtain approvals.
* The impact of the new EU General Data Protection Regulations remains poorly understood. Impact of BREXIT for UK studies is completely unknown.
* Most studies currently have very limited follow up (since device placement) and will not have accrued sufficient clinical events for meaningful analyses. Therefore, for most studies, linkages should possibly be re-visited in the future (e.g., 3-5 years’ time) and not prioritised at this moment. This will give ProPASS time to fine tune its sharing and data access process platforms and to carry out all necessary preparatory linkage work as outlined in “Linkage in ProPASS” section below.

1. **Background**

Data linking means bringing together two or more sources of information which relate to the same individual, event, institution or place. By combining the information it may be possible to identify relationships between factors which are not evident from the single sources.

The key advantage is that a linkage can transform an observational cohort into a long term prospective study at relatively low cost whilst reducing participant /study burden. While the use of sensitive data for research does create a confidentiality risk – and linked data have an increased risk – the empirical evidence suggests that this is an extremely low risk which can be managed effectively as research data is almost always de-identified as soon as is practical.

1. **Mechanics of linkage**

* When linking data, variables are typically split into ‘Identifying’ variables (for example, name, unique ID, medical insurance number) and variables of interest (age, gender, income, illness, mortality, occupation etc.)
* Direct identifiers (such as name, ID number) allow individuals to be identified exactly.
* Indirect identifiers only identify individuals in combination with other information.

There are four main types of data linkage:

1. *Exact/deterministic linking* - possible where a unique identifier is shared between two data sources. For example, in the UK, a National Health Service (NHS) number is used to link data across NHS medical records.
2. *Probabilistic data matching* - compares the identifying variables across two or more datasets to estimate the probability that two records relate to the same person.
3. *Statistical linking and data fusion* - records of two different individuals have been linked as if they refer to the same person (e.g., simulation models for policy evaluation).
4. *Multilevel linking* - not at the level of personal records. I.e., linking personal data with, for example, environmental/air pollution data.

There are five key stages of linkage:

1. Acquiring permission to link;
2. Agreeing the hosting protocol;
3. Acquiring the data;
4. Providing access to researchers;
5. Using linked data in research.

The first three stages needs agreement between multiple organisations. ‘Third party linking’ is common practise, where one organisation is given the identifying data only and creates an anonymous link field.

1. **Important institutional issues to consider during linkage.**
2. *Consent*

A person consenting for his or her confidential data to be linked and analysed is often referred to as the ‘gold standard’ gateway. There are a number of practical, ethical and statistical problems:

* it may be difficult or impractical to contact the participant;
* consent may lead to biased samples if those giving consent differ from those refusing it;
* use of data may identify family members, for example in DNA samples;
* gaining consent may be undesirable as it breaches confidentiality (for example, by revealing selection criteria).

1. *Competing jurisdictions*

Even if the legal framework is clearly defined, projects may suffer from needing the approval of multiple jurisdictions. This can be seen as a failure to distinguish between legal responsibility (to carry out due diligence on potential projects) and between gathering evidence (by accepting, for example, that another ethics committee is competent to carry out due diligence).

1. *Law versus custom*

Law is rarely a black-and-white issue; it needs interpretation in particular cases. However, most researchers are not specialists in law, and it is common for custom to be seen, over time, as law. This is most likely to occur where, in the absence of explicit legal statements, institutions are tasked with deciding the interpretation of the legal framework. Hence, research gateways can suffer from ‘regulatory capture’ by institutions keen to ensure that their interpretation of law prevails.

1. *Defining confidentiality*

Whilst legislation may use such terms as ‘confidential’ and ‘anonymised’, there is no legal definition. Instead it may be left open for a competent authority to determine, and/or reference to be taken to ‘reasonableness’. Hence, a key part of the legal framework is left open to human interpretation, and two organisations considering the confidentiality of a linked data source can come to different conclusions. This complicates any discussion on appropriate technical solutions.

1. **Linkage in ProPASS**
2. *Summary of survey results*

Our brief survey enquired about linkage possibilities across the current ProPASS studies. Our results suggest linked mortality will be the most readily available source of data across studies (see **Table 9**, **Appendix 5**). Hospital admission data are currently available in some studies and others have potential to collect such data. Cancer registry or other forms of linked data (e.g., environmental) appear less common.

We next enquired about the feasibility of sharing linked data sets across countries in order to create a harmonised pooled international data resource. Pooling linked data sets across countries appears feasible in the majority of cases (**Table 9**), although substantial variation may exist in the approach to obtain approvals.

1. *Proposed next steps and resource needed*

There was a clear lack of understanding on the impact of the new EU General Data Protection Regulations in relation to sharing linked data sets across countries. We know little about how the quality/validity of data linkage varies between studies. We need to better understand some of the key institutional issues of individual studies. In cases where anonymised data from individual studies cannot be easily exported, we would require technical expertise to undertake remote analyses and/or use group level estimates. Thus, this may require some discussion and agreed strategy about our general approach (i.e., using individual vs. group level estimate).

In conclusion, at this stage linkage would not seem to be a major priority. Particularly given that most studies currently have very limited follow up and will not have accrued sufficient clinical events for meaningful analysis until the next few years. Thus we would advise prioritising resource to other more pressing areas.

# **Workshop 6**

ProPASS Data Sharing Platforms

(And dealing with country specific legislation)

1. **Background**

The development of a data sharing platform for academic research involves a staged, multi-step process:

1. First, the Consortium (data platform) should have an explicit ethical and legal framework governing data sharing and policy documentation.
2. Secondly, Data platforms can facilitate information sharing, and prior agreement on the principles set above, but in the daily practice there is a huge need for expertise to handle the changing and needs in data management – knowledge management expertise.
3. Thirdly, we have the data sharing platform with infrastructure that enable data controllers and researchers to store, access and analyze sensitive data in controlled environments.

To determine and select the type of data sharing platforms one must be aware of data consortium components. (**Figure 2, Appendix 6.1**).

**Data components:**

* System design to transmit aggregate data to a central repository.
* Data-processing software.
* Selection of statistical techniques and parameters.

**Data sharing requires:**

* Expertise in several areas, including in-depth knowledge of the datasets to be linked
* Skills in the use of “sharing” software programs, includes merging of data sets
* Skills in statistical analysis and interpretation that comes from a multidisciplinary team including database managers and programmers
* Statisticians and data managements work collaboratively with researchers to resolve technical problems while keeping eye on the research question.
* Constant communication between site-level data providers, data coordinating centers, and principal investigators.

**Key components of successful data sharing**

Firstly, there is a need to think about how best to connect people for an effective sharing platform, because people share information with people they know. Secondly, there is a need to craft processes and tools aimed at different type of users and their needs. (**Figure 3, Appendix 6.1**).

**Key tooling/components for data sharing**

* **Migration:** Simple and secure transfer of data between studies or systems.
* **Storage:** Scalable option to build cost-effective collections.
* **Preservation:** Ensure integrity with data duplication (e.g. two sites, tapes)
* **Analytics:** Flexible tools for short cycle-times for cpu/gpu-based analysis
* **Result presentation:** To explain, exhibit and explore results (result servers)
* **Code repository:** To share code and software (e.g. git)
* **Knowledge preservation tools:** Continuous documentation (wiki and issues)
* **Communication tools:** Both internal/external (chats, twitter, web pages etc.)

1. **Data sharing platform options**

There are various types of sharing platforms; three types may be more relevant to ProPASS (decentralized, centralized and federated). (**Table 10, Appendix 6.2**)

The pros and cons are discussed below.

1. **Decentralized:** Local data store, no central storage. Is the simplest model, every data holder/study hosts its own data on its own server. External data requesters are allowed to establish user accounts on that server, perhaps with one access control. The requester then can view and analyze the data but cannot download a copy of the data to his/her own machine.

Data are largely disaggregated and lack technical and legal interoperability and, at most, may share a common index. The servers are in some way connected to each other.

Overall, this model is infeasible for widespread data sharing because it is prohibitively expensive and inefficient for multiple data holders to handle access control, data provision, and user account services. Moreover, data requesters wishing to query multiple databases—for example, to carry out a meta-analysis—must establish multiple user accounts and navigate multiple access policies and procedures. Even after obtaining access to multiple data sets, the requester could not merge them.

1. **Centralized:** The opposite of having each data holder maintain its own server is to collect all data into one central database. All data are aggregated in a single, centrally managed repository. There are multiple issues with data exportation (both accelerometry and linkage data), as identified in the Pro PASS Survey (**Table 11, Table 12, Table 13, Appendix 7**).
2. **Federated** (Includes a semi-centralized approach):

The federated query model combines the approach of bringing the query to the data with federated databases. Databases are federated when independent geographically dispersed databases are networked in such a way that they can respond to queries as if all the data were in a single virtual database. Thus, when data requesters submit a query to a federated query service, that query is routed to all databases participating in the federation. The provider of the query service may or may not be the “trusted intermediary” that adjudicates access control requests. Federated query services can be purchased as a standalone technical service. Data holders maintain full control over their data, and neither the data requester nor the query service provider ever has direct access to the data. Federated query systems can protect against invasions of privacy

**Semi-distributed/separated:** Data are distributed and separately maintained, sometimes across national borders, but may be interconnected by a central access portal, may share other technical service components, and may utilize a common data-exchange format.

**Fully distributed:** Data are maintained locally and are not technically integrated, but share a common legal and policy framework that allows access on uniform terms and conditions (legal interoperability).

1. **Possible solutions**

See **Figure 4** (**Appendix 6.2**).

The systems ProPASS choose, need to meet some essential criteria such as:

* Data controllers and researchers must be able to store, access and analyze sensitive data in controlled environments.
* Systems must be adaptive and flexible, so they can be adjusted and grown to meet our changing needs throughout the project period.
* Ensure interoperability and deployment flexibility between clouds or within our own infrastructure.

HUNT Cloud are organized in digital labs:

* The Labsare **logically separated** and controlled.
* A consortium can establish **several labs.**
* **Users** can be authorized to one or several labs.
* Establishment require a valid **data processor agreement** with the institution that control the data that will be held in the lab.
* There is viable **cloud community** with experience in large scale genetics consortiums.

See **Figure 5** (**Appendix 6.2**).

HUNT Cloud delivers digital infrastructure that enable data controllers and researchers to store, access and analyze sensitive data in controlled environments:

* **Migration:** Internal and external solutions.
* **Storage:** Low cost and scalable on commodity disks.
* **Preservation:** Encrypted tapes sent to data controllers for requested data.
* **Analytics:** Flexible CPU/GPU-based resources for qc/analysis of big data.
* **Access control:** Customizable security tiers compliant with e.g. GDPR.

HUNT Cloud is committed to international excellence in privacy and information security!

They are the first academic research cloud in Norway with third-party certified management systems both for quality and information security (ISO 27001 and ISO 9001).

**Points for discussion**

* What are the essential tools/components for data sharing in ProPASS?
* What type of model and solution is preferable for ProPASS?
* How can the data sharing and access platform development be divided into logical stages/work packages?
* What are the resource (time, monetary costs, expertise) requirements for each work package?

**References for this section:** (Bergeron, et al., 2018)**,** (Carter, et al., 2016)**,** (Conteras & Reichman, 2015)**.** (Doiron, et al., 2013)**,** (Doiron, et al., 2017)**,** (Firnkorn, et al., 2015)**,** (Fortier, et al., 2015)**,** (Kaye & Hawkins, 2014)**,** (Lakerveld, et al., 2017).

# **References**

Aguilar-Farias, N., Brown, W. J., Olds, T. S. & Peeters, G., 2015. Validity of self-report methods for measuring sedentary behaviour in older adults. *Journal of Science and Medicine,* 18(6), pp. 662-666.

Aguilar-Farias, N., Brown, W. J. & Peeters, G., 2014. ActiGraph GT3X+ cut-points for identifying sedentary behaviour in older adults in free-living environment. *Journal of Science and Medicine in Sport,* 17(3), pp. 293-299.

Atkin, A. J. et al., 2017. Harmonising data on the correlates of physical activity and sedentary behaviour in young people: Methods and lessons learnt from the international Children’s Accelerometry database (ICAD). *BMC,* Volume 14, p. 174.

Barreira, T. et al., 2015. Free-living activity counts-derived breaks in sedentary time: Are they real transitions from sitting to standing?. *Gait & Posture,* 42(1), pp. 70-72.

Bergeron, J. et al., 2018. Fostering population-based cohort data discovery: The Maelstrom Research cataloguing toolkit. *PLoS One,* 13(7), p. e0200926.

Carter, K. W., Francis, R. W., Bresnahan, M. & Gissler, M., 2016. ViPAR: a software platform for the Virtual Pooling and Analysis of Research Data. *International Journal of Epidemiology,* 45(2), pp. 408-416.

Clark, B. K. et al., 2018. Using Bluetooth proximity sensing to determine where office workers spend time at work. *PLoS One,* 13(3), p. e0193971.

Clemes, S. et al., 2012. Validity of the ActiGraph GT3X+ inclinometer and different counts per minutes cut-points for the assessment of sedentary behaviour. *Journal of Science and Medicine in Sport,* 15(1), p. S68.

Cochrane, 2018. *Cochrane Methods: What is a PMA?.* [Online]   
Available at: https://methods.cochrane.org/pma/what-pma-0  
[Accessed September 2018].

Conteras, J. L. & Reichman, J. H., 2015. DATA ACCESS. Sharing by design: Data and decentralized commons. *Science,* 350(6266), pp. 1312-1314.

Cooper, K. et al., 2017. Accuracy of physical activity recognition from a wrist-worn sensor. *Physiotherapy,* 103(Supplement 1), p. e47.

Denkinger, M., 2012. Accelerometer-based physical activity in a large observational cohort (ActiFE-Ulm) - Associations with multimeidication and falls. *Internist,* Volume 1, pp. 52-53.

Deshmukh, P., Russell, C., Lucarino, L. & Robinovitch, S., 2012. *Enhancing clinical measures of postural stability with wearable sensors.* s.l., s.n., pp. 4521-4.

Doiron, D. et al., 2013. Data harmonisation and federated analysis of population-based studies: the BioSHaRE project. *Emerging Themes in Epidemiology,* 10(1), p. 12.

Doiron, D. et al., 2017. Software Application Profile: Opal and Mica: open-source software solutions for epidemiological data management, harmonization and dissemination. *International Journal of Epidemiology,* 46(5), pp. 1372-1378.

Edwardson, C. L. et al., 2017. Considerations when using the activPAL monitor in field-based research with adult populations. *Journal of Sport and Health Science,* Volume 6, pp. 163-178.

Ekelund, U. et al., 2012. Moderate to Vigorous Physical Activity and Sedentary Time and Cardiometabolic Risk Factors in Children and Adolescents. *JAMA,* 307(7), p. 704–712.

Firnkorn, D. et al., 2015. A Generic Data Harmonization Process for Cross-linked Research and Network Interaction. Construction and Application for the Lung Cancer Phenotype Database of the German Center for Lung Research. *Methods Inf Med,* 54(5), pp. 455-460.

Florencio, T. et al., 2015. Weight gain and reduced energy expenditure in low-income Brazilian women living in slums: a 4-year follow-up study. *British Journal of Nutrition,* Volume 114, pp. 462-471.

Fortier, L. P. et al., 2015. The RECITE Study: A Canadian Prospective, Multicenter Study of the Incidence and Severity of Residual Neuromuscular Blockade. *Anesthesia & Analgesia,* 121(2), pp. 366-372.

Fortune, E., Lugade, V., Amin, S. & Kaufman, K., 2015. Step detection using multi- versus single tri-axial accelerometer-based systems. *Physiological Measurement,* Volume 36, p. 2519.

Gao, Y., Cronin, N., Nevala, N. & Finni, T., 2017. Validity of long-term and short-term recall of occupational sitting time in Finnish and Chinese office workers. *Journal of Sport and Health Science,* p. 108.

Griffith, L. E. et al., 2016. Comparison of Standardization Methods for the Harmonization of Phenotype Data: An Application to Cognitive Measures. *American Journal of Epidemiology,* 184(10), pp. 770-778.

Griffith, L. et al., 2013. *Harmonization of Cognitive Measures in Individual Participant Data and Aggregate Data Meta-Analysis,* Rockville: Agency for Healthcare Research and Quality (US).

Gupta, N. et al., 2016a. Are temporal patterns of sitting associated with obesity among blue-collar workers? A cross sectional study using accelerometers. *BMC Public Health,* Volume 16, p. 148.

Gupta, N. et al., 2016b. What Is the Effect on Obesity Indicators from Replacing Prolonged Sedentary Time with Brief Sedentary Bouts, Standing and Different Types of Physical Activity during Working Days? A Cross-Sectional Accelerometer-Based Study among Blue-Collar Workers. *PLoS One,* 11(5), p. e0154935.

Gupta, N., Heiden, M., Mathiassen, S. & Holtermann, A., 2018. Is self-reported time spent sedentary and in physical activity differentially biased by age, gender, body mass index, and low-back pain?. *Scandinavian Journal of Work, Environment and Health,* 44(2), pp. 163-170.

Hallman, D. M., Jorgensen, M. B. & Holtermann, A., 2017. On the health paradox of occupational and leisure-time physical activity using objective measurements: Effects on autonomic imbalance. *PLoS One,* 12(5), p. e0177042.

Hallman, D. M. et al., 2015b. Differences between work and leisure in temporal patterns of objectively measured physical activity among blue-collar workers. *BMC Public Health,* Volume 15, p. 976.

Hallman, D. M. et al., 2016. Temporal patterns of sitting at work are associated with neck–shoulder pain in blue-collar workers: a cross-sectional analysis of accelerometer data in the DPHACTO study. *International Archives of Occupational and Environmental Health,* 89(5), pp. 823-833.

Hallman, D. M. et al., 2015c. Prolonged Sitting is Associated with Attenuated Heart Rate Variability during Sleep in Blue-Collar Workers. *Int J Environ Res Public Health,* 12(11), pp. 14811-14827.

Hallmann, D. M., Gupta, N., Mathiassen, S. E. & Holtermann, A., 2015a. Association between objectively measured sitting time and neck–shoulder pain among blue‑collar workers. *Int Arch Occup Environ Health,* Volume 88, pp. 1031-1042.

Ioannidis, J., 2017. Next-generation systematic reviews: prospective meta-analysis, individual-level data, networks and umbrella reviews. *British Journal of Sports Medicine,* 51(20), pp. 1456-1458.

Kaye, J. & Hawkins, N., 2014. Data sharing policy design for consortia: challenges for sustainability. *Genome Med,* 6(1), p. 4.

Klenk, J. et al., 2013. Association of objectively measured physical activity with established and novel cardiovascular biomarkers in elderly subjects: every step counts. *J Epidemiol Community Health,* Volume 67, pp. 194-197.

Klenk, J. et al., 2015. Objectively measured physical activity and vitamin D status in older people from Germany. *J Epidemiol Community Health,* Volume 69, pp. 388-392.

Kloster, S. et al., 2017. How Does Definition of Minimum Break Length Affect Objective Measures of Sitting Outcomes Among Office Workers?. *Journal of Physical Activity & Health,* 14(1), pp. 8-12.

Koster, A. et al., 2016. Comparison of Sedentary Estimates between activPAL and Hip- and Wrist-worn ActiGraph. *Med Sci Sports Exerc,* 48(8), pp. 1514-1522.

Kurita, S. et al., 2017. Comparability of activity monitors used in Asian and Western-country studies for assessing free-living sedentary behaviour. *PLoS ONE,* 12(10), p. e0186523.

Lakerveld, J. et al., 2017. Identifying and sharing data for secondary data analysis of physical activity, sedentary behaviour and their determinants across the life course in Europe: general principles and an example from DEDIPAC. *BMJ Open ,* p. e017489.

Liu, S. H. & Chang, Y., 2009. Using accelerometers for physical activity actions recognition by a neural fuzzy network. *Telemedicine Journal & E-Health,* 15(9), pp. 867-876.

Lyons, G. M. et al., 2005. A description of an accelerometer-based mobility monitoring technique. *Medical Engineering & Physics,* 27(6), pp. 497-504.

Mannini, A. & Sabatini, A., 2011. *On-line classification of human activity and estimation of walk-run speed from acceleration data using support vector machines.* s.l., s.n., pp. 3302-3305.

Matthews, C. E., Hagströmer, M., Pober, D. & Bowles, H. R., 2012. Best practices for using physical activity monitors in population-based research. *Med Sci Sports Exerc,* 44(1 Suppl 1), pp. S68-76.

Moher, D., Liberati, A., Tetzlaff, J. & Altman, D. G., 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMAS Statement. *PLoS Medicine (OPEN ACCESS),* 6(7), p. e1000097.

National Institute of Health, 2014. *Study Quality Assessment Tools.* [Online]   
Available at: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools  
[Accessed 26 July 2018].

Public Health Research Data Forum, March 2015. *Enabling Data Linkage to Maximise the Value of Public Health Research Data: full report,* s.l.: Commissioned by the Wellcome Trust.

Pulakka, A. et al., 2018. Association Between Employment Status and Objectively Measured Physical Activity and Sedentary Behavior—The Maastricht Study. *Journal of Occupational and Environmental Medicine,* 60(4), pp. 309-315.

Riou, M., Rioux, F., Lamothe, G. & Doucet, E., 2015. Validation and Reliability of a Classification Method to Measure the Time Spent Performing Different Activities. *PLoS ONE,* 10(6), p. e0128299.

Schneller, M. et al., 2015. Validation of Five Minimally Obstructive Methods to Estimate Physical Activity Energy Expenditure in Young Adults in Semi-Standardized Settings. *Sensors,* Volume 15, pp. 6133-6151.

Skarpsno, E. S. et al., 2018. Objectively measured occupational and leisure-time physical activity: cross-sectional associations with sleep problems. *Scandinavian Journal of Work, Environment & Health,* 44(2), pp. 202-211.

Skotte, J. et al., 2014. Detection of Physical Activity Types Using Triaxial Accelerometers. *Journal of Physical Activity & Health,* 11(1), pp. 76-84.

Steeves, J. A. et al., 2015. Ability of thigh-worn ActiGraph and activPAL monitors to classify posture and motion. *Medicine and Science in Sports and Exercise,* 47(5), pp. 952-959.

Stemland, I. et al., 2015. Validity of the Acti4 method for detection of physical activity types in free-living settings: comparison with video analysis. 58(6), pp. 953-965.

Stroup, D. F. et al., 2000. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *Journal of American Medical Association,* 19 April, 283(15), pp. 2008-12.

UCL, 2018. *Cohort and Longitudinal Studies Enhancement Resources (CLOSER).* [Online]   
Available at: http://www.ucl.ac.uk/ioe/departments-centres/centres/cohort-and-longitudinal-studies-enhancement-resources  
[Accessed October 2018].

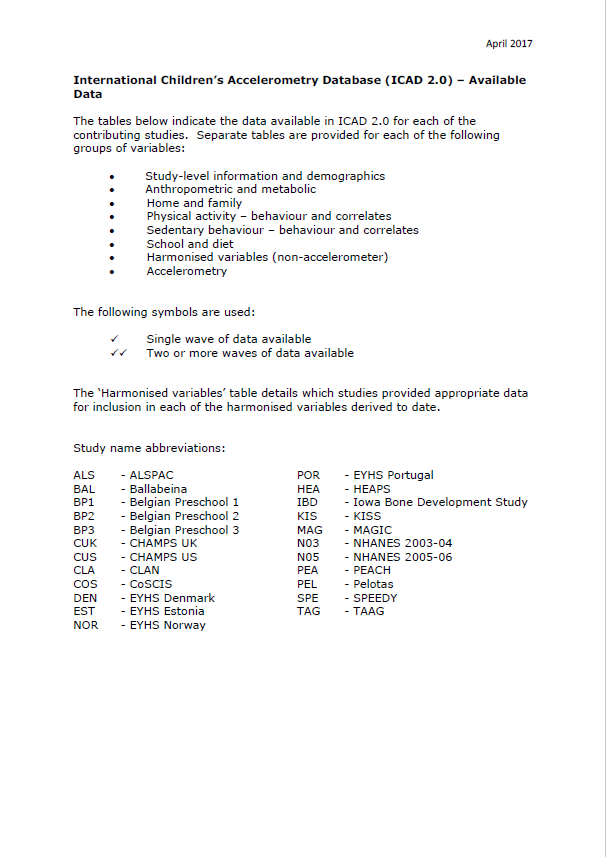
von Elm, E. et al., 2007. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Annals of Internal Medicine,* pp. 147(8), 573.

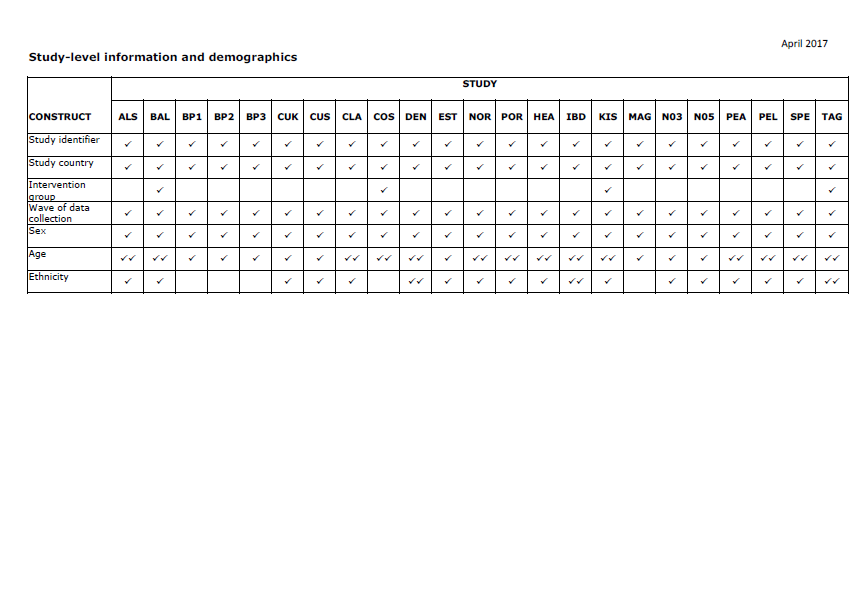
Wells, G. A. et al., 2009. *The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomized studies in meta-analyses.* [Online]   
Available at: http://www.ohri.ca/programs/clinical\_epidemiology/oxford.asp  
[Accessed 26 July 2018].

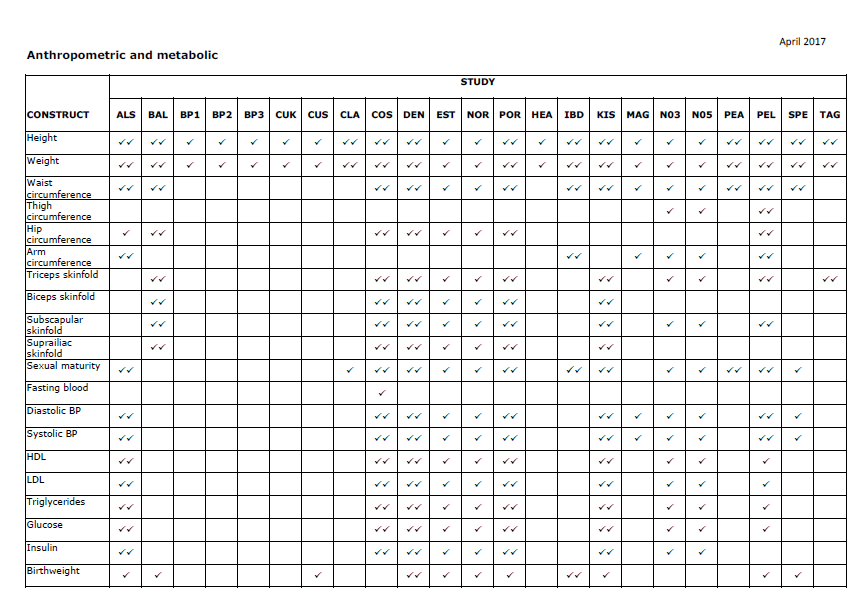
Wullems, J. A. et al., 2017. Performance of thigh-mounted triaxial accelerometer algorithms in objective quantification of sedentary behaviour and physical activity in older adults. *PLoS ONE,* 12(11), p. e0188215.

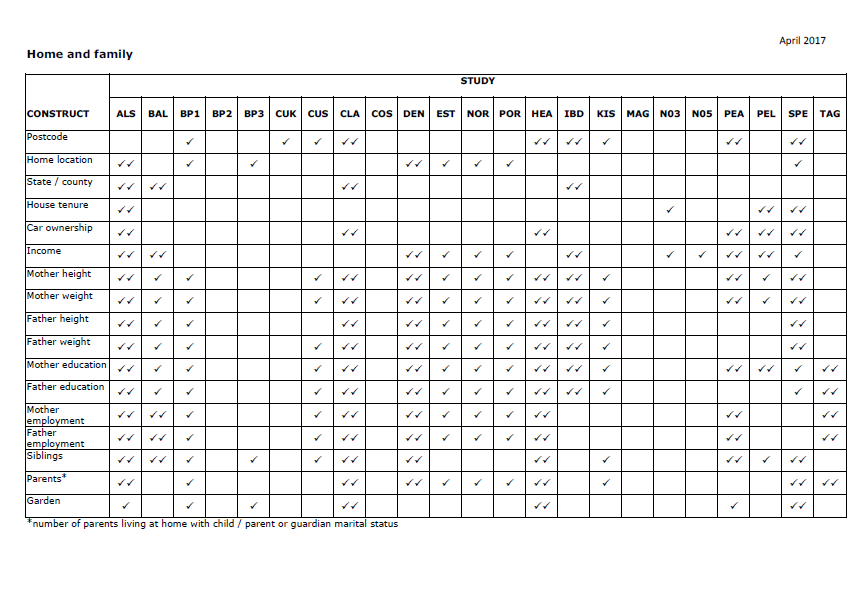
# **Appendix 1: Workshop 1**

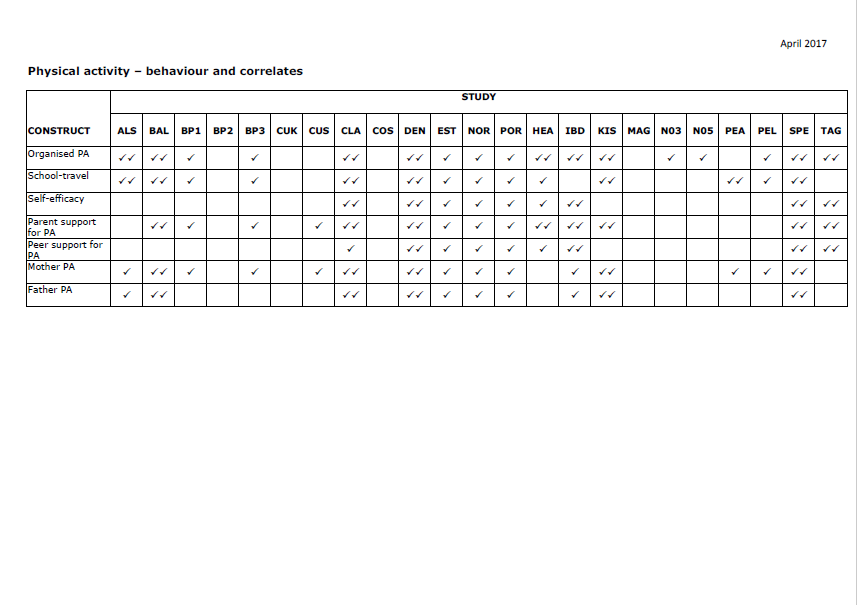
## Appendix 1.1: ICAD expanded list of harmonized variables

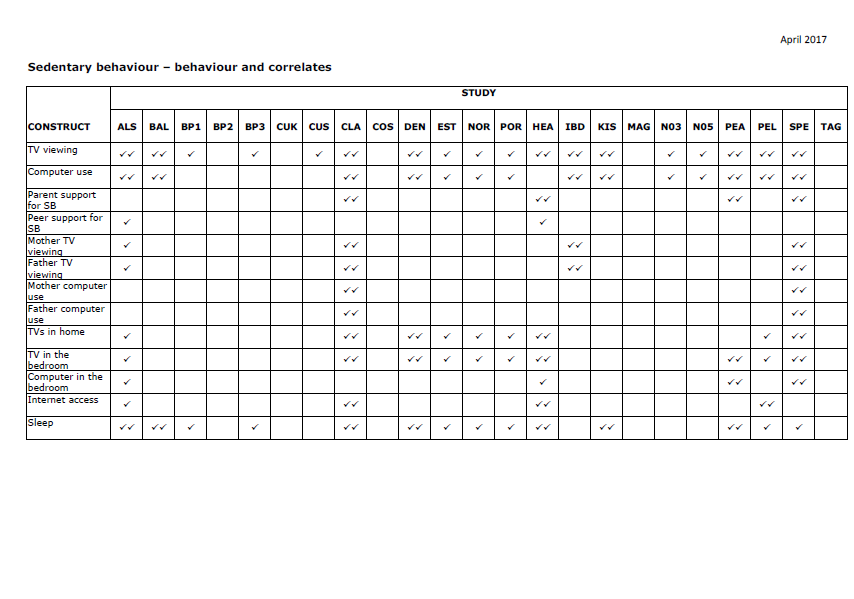


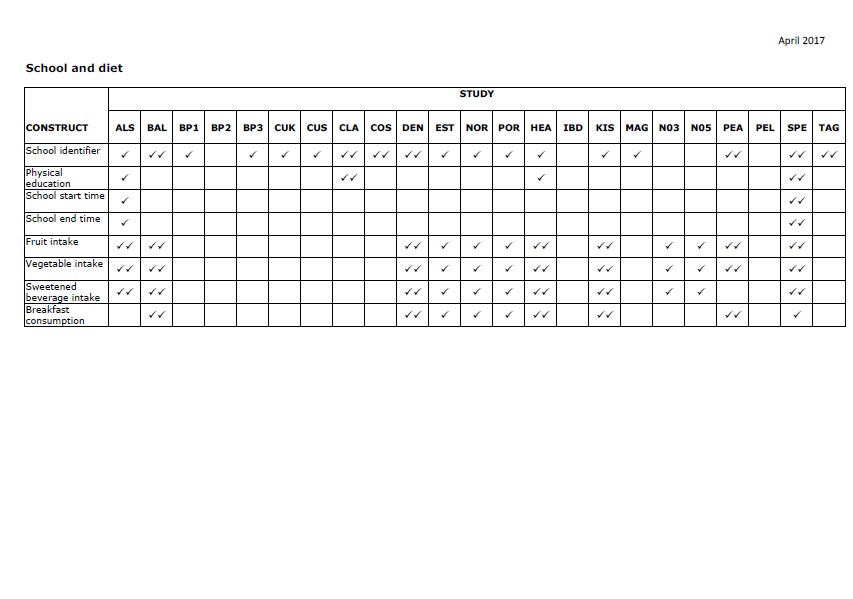


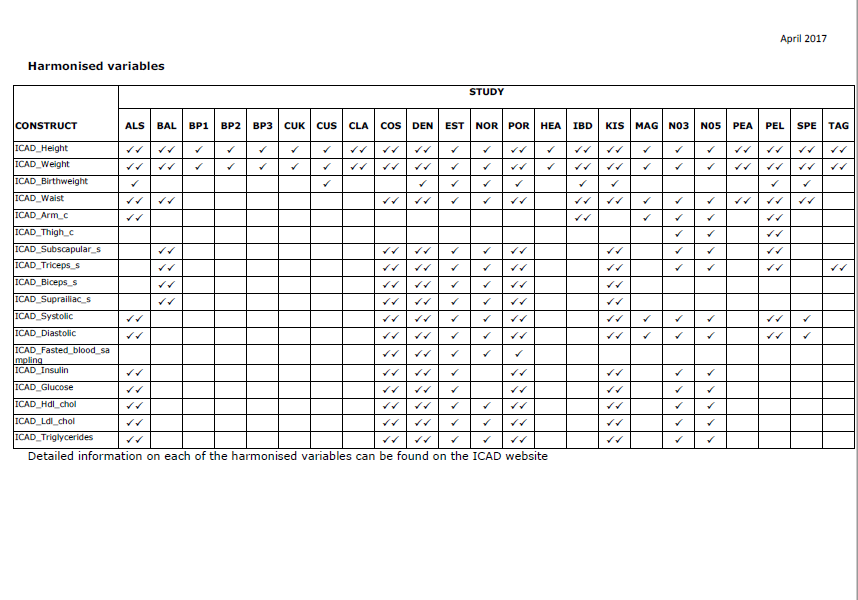


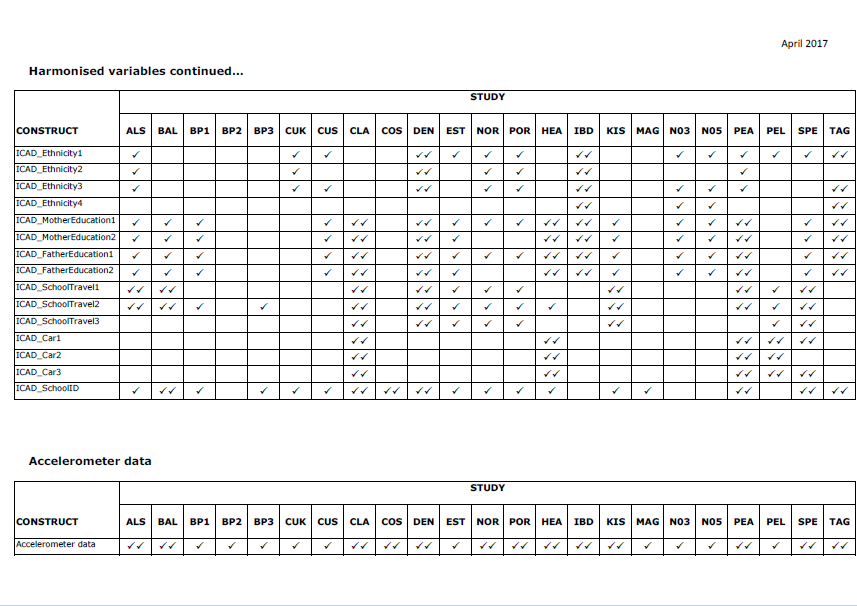












## Appendix 1.2: Protocol Overview

**Tri-axial thigh-worn accelerometers for physical activity, posture, and sleep behaviour quantification in adults: A systematic scoping review of observational studies (PROTOCOL)**

**AIM & OBJECTIVES**

We aim to systematically identify and describe all observational studies that used 24 hour data of tri-axial thigh-worn accelerometry to track physical activity, posture, and sleep behaviour quantification data of free-living adult participants in population-based or community settings with the objectives to:

1. Identify all observational cohorts that can be potentially included in our own consortium project;
2. Assemble information to describe each study’s population and samples, recruitment methods, sample demographic characteristics, study settings, and summarise the groups of variables used (specific exposures, confounders, mediators, moderators, outcomes) to target;
3. Review and summarise accelerometry data collection methods, accelerometry data processing methods (including software used), and procedural and data-related characteristics.

**METHODS**

This review will be conducted and reported according to the MOOSE reporting standards (Meta-analysis Of Observational Studies in Epidemiology: a proposal for reporting) (Stroup, et al., 2000), along with relevant elements of the PRISMA reporting standards (Moher, et al., 2009) such as study flow chart. The MOOSE items checklist and flow chart will be presented in tabular format.

**Eligibility Criteria for Included Studies**

1. **Types of studies**
   1. ***Designs***

We will include all observational study designs such as prospective cohort, retrospective cohort (historical cohort), and cross-sectional study designs.

We require full text publication. This includes published pilot studies, or published study protocols.

We will include published abstracts that contains information to clearly suffice our inclusion criteria and, also contain sufficient information to extract data on at least our primary outcome (characteristics of the study). It the abstract suffices the inclusion criteria but does not provide enough information for our primary outcome, we will attempt to obtain information from the authors. Otherwise the abstract will sit in the awaiting classification list.

We anticipate that we may find multiple papers belonging to a ‘parent study’. We plan to include all sub-papers of mother studies and count them as individual studies with separate data extraction, as different outcomes may be reported between these multiple papers. We will record which ‘mother study’ these are linked to.

We do not require a minimum number of participants per study to be included in this review as we are interested in their recruitment, methods and analysis procedures.

We will exclude interventional studies of any type.

* 1. ***Measurements***

We will include studies where participants wear thigh-worn tri-axial accelerometers, for any number of days that utilized 24 hour activity data monitoring protocols. This includes studies where participants wear alternative accelerometers, in addition to the thigh placement. We will report only on the separate thigh-worn data, if available.

1. **Participants**
   1. ***Inclusion criteria***

* Population or community-based cohorts of free-living adults;
* Aged 18 years and older;
* Must be recruited through a community or population setting such as a GP or hospital/clinical setting, where the participants themselves are not living in a restricted or controlled setting;
* Examples include, but are not limited to: cohorts of cardiac rehabilitation patients, hypertension patients, and healthy individuals.
  1. ***Exclusion criteria***
* Institutionalised participants e.g. people living in care homes or hospitalised patient cohorts;
* Specialised clinical cohorts currently undergoing (or perioperative) major treatments or major surgery, must be >30 days postoperative;
* Certain therapies which would hinder physical activity levels such as: oncological treatment/therapy, renal dialysis, participants with HIV, Hepatitis C, or any type of arthritis;
* If studies include some participants under 18 years (<20%), we may consider to include on a case-by-case basis so long as the participant is close within adolescents/teenage years.

1. **Types of outcome measures**

We will not exclude studies based on the availability of reported outcome data.

* 1. ***Primary outcomes***

1. Characteristics of studies (design, recruitment, inclusion/exclusion criteria, baseline characteristics, outcomes, size, setting, and location).
2. ***Secondary outcomes***
3. Primary and secondary outcomes reported by the study: divided into physical activity and non-physical activity variables. We are particularly interested in and will take note of whether they reported: physical activity, posture, estimates of sleep, health status of the participants (such as comorbidities, mortality) and if they reported on adverse effects;
4. Accelerometry data collection processing models used, and data analysis;
5. Data sharing methods and platforms, if information is available.

**Data Collection and Analysis**

1. **Selection of studies**

Two reviewers will independently screen the titles and abstracts for eligibility. We will obtain the full texts of potentially eligible studies and independently assess these for inclusion or exclusion. In the case of any disagreement, we will discuss and/or resolve with a third author if necessary.

We will compare and report our search findings according to PRISMA standards (Moher, et al., 2009) in a flow diagram.

1. **Data extraction and management**

Two authors will independently extract the information of included studies into a pre-agreed data extraction template in table format.

We will extract any available information on:

* Study name and citation; if the study is an individual publication or linked to a mother study;
* Our primary and secondary outcomes

1. Characteristics of studies (design, recruitment, inclusion/exclusion criteria, baseline characteristics, outcomes, size, location).
2. Primary and secondary outcomes reported by the study: divided into physical activity and non-physical activity variables. We are particularly interested in and will take note of whether they reported: physical activity, posture, estimates of sleep, health status of the participants (such as comorbidities, mortality) and if they reported on adverse effects;
3. Accelerometry data collection processing models used, and data analysis;
4. Data sharing methods and platforms, if information is available.

We will also record on the form for each study:

* Quality assessment (using NOS or NIH scales applicable to the study design)
* Other important details to note about the study related to our review
  + STROBE, disclosures and funding, ethics, other issues that might arise

1. **Quality assessment**

Two authors will independently assess and rate the quality of the included studies using one of two available scales:

* The Newcastle-Ottawa Scale (NOS) (Wells, et al., 2009) will be used for our included cohort and case-control studies and;
* The NIH Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies (National Institute of Health, 2014) will be used for our included cross-sectional studies - as NOS does not cover cross-sectional studies.

In addition to these scales, we will further assess the quality of the studies by checking the following:

* Did the authors adhere to the STROBE reporting standards (von Elm, et al., 2007) in their manuscript or full text?
* Did the authors report on/gain written informed consent and ethics?
* Did the authors report on disclosures and funding?
* Any other issues that might arise we wish to comment on

1. **Quantitative assessment**

We will quantitatively report the basic primary and secondary outcome data. Further details of these will be narratively reported.

We do not plan any statistical analyses or meta-analyses for this scoping review.

1. **Reporting**

Two review authors will compare and collate data extraction. In the case of any disagreement, we will resolve with a third author if necessary.

We plan to present our findings in narrative form as well as present our results in visual and tabular format:

* Figure of Study Flowchart
* Primary and Secondary outcomes table of results
* Quality Assessment checklist tables
* STROBE checklist table

## Appendix 1.3: Full list of included studies for Literature Review #1

|  |  |  |
| --- | --- | --- |
|  | **Citation** | **Parent Study** |
|  | Aguilar-Farias, N., et al. (2014). "ActiGraph GT3X+ cut-points for identifying sedentary behaviour in older adults in free-living environments." Journal of Science & Medicine in Sport 17(3): 293-299. | - |
|  | Barreira, T. V., et al. (2015). "Free-living activity counts-derived breaks in sedentary time: Are they real transitions from sitting to standing?" Gait & Posture 42(1): 70-72. | - |
|  | Bellettiere, J. (2017). "Associations of sitting accumulation patterns with cardio-metabolic risk biomarkers in Australian adults." PLoS ONE [Electronic Resource]. | AusDiab |
|  | Dall, P. M., et al. (2017). "The influence of dog ownership on objective measures of free-living physical activity and sedentary behaviour in community-dwelling older adults: a longitudinal case-controlled study." BMC Public Health 17(1): 496. | - |
|  | de Rooij, B. H., et al. (2016). "Physical Activity and Sedentary Behaviours in Metabolically Healthy versus Unhealthy Obese and Non-Obese Individuals - The Maastricht Study." PLoS ONE [Electronic Resource] 11(5): e0154358. | Maastricht Study |
|  | Denkinger, M. D., et al. (2010). "Accelerometer-based physical activity in a large observational cohort--study protocol and design of the activity and function of the elderly in Ulm (ActiFE Ulm) study." BMC Geriatrics 10: 50. | ActiFE |
|  | Fisher, A., et al. (2018). "Associations between the objectively measured office environment and workplace step count and sitting time: Cross-sectional analyses from the active buildings study." International Journal of Environmental Research and Public Health 15 (6) (no pagination)(1135). | Active Buildings |
|  | Florencio, T. M., et al. (2015). "Weight gain and reduced energy expenditure in low-income Brazilian women living in slums: a 4-year follow-up study." British Journal of Nutrition 114(3): 462-471. | - |
|  | Granat, M., et al. (2015). "Quantifying the cadence of free-living walking using event-based analysis." Gait & Posture 42(1): 85-90. | - |
|  | Gupta N, Christiansen CS, Hallman DM, Korshøj M, Carneiro IG, Holtermann A (2015) Is Objectively Measured Sitting Time Associated with Low Back Pain? A Cross-Sectional Investigation in the NOMAD study. PLoS ONE 10(3): e0121159. doi:10.1371/journal.pone.0121159 | NOMAD |
|  | Gupta, N., et al. (2016). "Are temporal patterns of sitting associated with obesity among blue-collar workers? A cross sectional study using accelerometers." BMC Public Health 16: 148. | NOMAD |
|  | Gupta, N., et al. (2016). "What Is the Effect on Obesity Indicators from Replacing Prolonged Sedentary Time with Brief Sedentary Bouts, Standing and Different Types of Physical Activity during Working Days? A Cross-Sectional Accelerometer-Based Study among Blue-Collar Workers." PLoS ONE [Electronic Resource] 11(5): e0154935. | DPHACTO |
|  | Gupta, N., et al. (2018). "Is self-reported time spent sedentary and in physical activity differentially biased by age, gender, body mass index, and low-back pain?" Scandinavian Journal of Work, Environment and Health 44(2): 163-170. | NOMAD |
|  | Hallman, D. M., et al. (2017). "On the health paradox of occupational and leisure-time physical activity using objective measurements: Effects on autonomic imbalance." PLoS ONE [Electronic Resource] 12(5): e0177042. | DPHACTO |
|  | Hallman, D. M., et al. (2017). "Objectively measured physical activity and 12-month trajectories of neck-shoulder pain in workers: A prospective study in DPHACTO." Scandinavian Journal of Public Health 45(3): 288-298. | DPHACTO |
|  | Hallman, D. M., et al. (2016). "Is prolonged sitting at work associated with the time course of neck-shoulder pain? A prospective study in Danish blue-collar workers." BMJ Open 6(11): e012689. | DPHACTO |
|  | Hallman, D. M., et al. (2016). "Temporal patterns of sitting at work are associated with neck-shoulder pain in blue-collar workers: a cross-sectional analysis of accelerometer data in the DPHACTO study." International archives of occupational and environmental health 89(5): 823-833. | NOMAD |
|  | Hallman, D. M., et al. (2015). "Association between objectively measured sitting time and neck-shoulder pain among blue-collar workers." International Archives of Occupational & Environmental Health 88(8): 1031-1042. | NOMAD |
|  | Hallman, D. M., et al. (2015). "Differences between work and leisure in temporal patterns of objectively measured physical activity among blue-collar workers." BMC Public Health 15: 976. | DPHACTO |
|  | Hallman, D. M., et al. (2015). "Prolonged Sitting is Associated with Attenuated Heart Rate Variability during Sleep in Blue-Collar Workers." International Journal of Environmental Research & Public Health [Electronic Resource] 12(11): 14811-14827. | NOMAD |
|  | Hulsegge, G., et al. (2017). "Shift workers have similar leisure-time physical activity levels as day workers but are more sedentary at work." Scandinavian Journal of Work, Environment & Health 43(2): 127-135. | NOMAD |
|  | JHPM, V. D. V., et al. (2017). "Sedentary Behavior, Physical Activity, and Fitness-The Maastricht Study." Medicine & Science in Sports & Exercise 49(8): 1583-1591. | Maastricht Study |
|  | Klenk, J., et al. (2013). "Association of objectively measured physical activity with established and novel cardiovascular biomarkers in elderly subjects: every step counts." Journal of Epidemiology & Community Health 67(2): 194-197. | ActiFE |
|  | Klenk, J., et al. (2015). "Physical Activity and Different Concepts of Fall Risk Estimation in Older People--Results of the ActiFE-Ulm Study." PLoS ONE [Electronic Resource] 10(6): e0129098. | ActiFE |
|  | Klenk, J., et al. (2015). "Objectively measured physical activity and vitamin D status in older people from Germany." Journal of Epidemiology & Community Health 69(4): 388-392. | ActiFE |
|  | Lagersted-Olsen, J., et al. (2014). "Comparison of objectively measured and self-reported time spent sitting." International Journal of Sports Medicine 35(6): 534-540. | DPHACTO |
|  | Lagersted-Olsen, J., et al. (2016). "Does objectively measured daily duration of forward bending predict development and aggravation of low-back pain? A prospective study." Scandinavian Journal of Work, Environment & Health 42(6): 528-537. | DPHACTO |
|  | Martens, R. J. H., et al. (2018). "Amount and pattern of physical activity and sedentary behavior are associated with kidney function and kidney damage: The Maastricht Study." PLoS ONE [Electronic Resource] 13 (4) (no pagination)(e0195306). | Maastricht Study |
|  | Matsuo, T., et al. (2016). "Percentage-Method Improves Properties of Workers' Sitting- and Walking-Time Questionnaire." Journal of Epidemiology 26(8): 405-412. | - |
|  | Mazzotta, M. A., et al. (2018). "Usage of Sit-Stand Workstations and Associations between Work and Nonwork Sitting Time: An Observational Study." Journal of Occupational and Environmental Medicine 60(5): e268-e272. | - |
|  | Pulakka, A., et al. (2018). "Association between Employment Status and Objectively Measured Physical Activity and Sedentary Behavior-The Maastricht Study." Journal of Occupational and Environmental Medicine 60(4): 309-315. | Maastricht Study |
|  | Rasmussen, C. L., et al. (2018). "Does physically demanding work hinder a physically active lifestyle in low socioeconomic workers? A compositional data analysis based on accelerometer data." International Journal of Environmental Research and Public Health 15 (7) (no pagination)(1306). | DPHACTO |
|  | Sawyer, A., et al. (2017). "Perceived office environments and occupational physical activity in office-based workers." Occupational Medicine (Oxford) 67(4): 260-267. | Active Buildings |
|  | Shaw, R. J., et al. (2017). "The Influence of Neighbourhoods and the Social Environment on Sedentary Behaviour in Older Adults in Three Prospective Cohorts." International Journal of Environmental Research & Public Health [Electronic Resource] 14(6): 24. | Seniors USP (Understanding Sedentary Patterns) |
|  | Skarpsno, E. S., et al. (2018). "Objectively measured occupational and leisure-time physical activity: Cross-sectional associations with sleep problems." Scandinavian Journal of Work, Environment and Health 44(2): 202-211. | DPHACTO |
|  | Smith L, Ucci M, Marmot A, et al. Active buildings: modelling physical activity and movement in office buildings. An observational study protocol. BMJ Open 2013;3:e004103. doi:10.1136/bmjopen-2013-004103 | Active Buildings |
|  | Smith, L., et al. (2015). "Weekday and weekend patterns of objectively measured sitting, standing, and stepping in a sample of office-based workers: the active buildings study." BMC Public Health 15: 9. | Active Buildings |
|  | Smith, L., et al. (2018). "Occupational physical activity habits of UK office workers: Cross-sectional data from the active buildings study." International Journal of Environmental Research and Public Health 15 (6) (no pagination)(1214). | Active Buildings |
|  | Smith, L., et al. (2014). "The association between objectively measured sitting and standing with body composition: A pilot study using MRI." BMJ Open 4 (6) (no pagination)(e005476). | - |
|  | Spinney, R., et al. (2015). "Indoor Tracking to Understand Physical Activity and Sedentary Behaviour: Exploratory Study in UK Office Buildings." PLoS ONE [Electronic Resource] 10(5): e0127688. | Active Buildings |
|  | Straker, L., et al. (2013). "Sit-stand desks in call centres: associations of use and ergonomics awareness with sedentary behavior." Applied Ergonomics 44(4): 517-522. | - |
|  | van der Berg, J. D., et al. (2016). "Associations of total amount and patterns of sedentary behaviour with type 2 diabetes and the metabolic syndrome: The Maastricht Study." Diabetologia 59(4): 709-718. | Maastricht Study |
|  | Wick, K., et al. (2016). "Deviation between self-reported and measured occupational physical activity levels in office employees: effects of age and body composition." International Archives of Occupational & Environmental Health 89(4): 575-582 | - |

## Appendix 1.4: Table of non-accelerometer variables

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 1**: Non-accelerometry variables available in 7 randomly selected studies of thigh-worn accelerometry. | | | | | | | |
| **Variables Group/ Variable Family** | **(1)**  **Bellettiere 2017** | **(2)**  **de Rooij 2016** | **(3)**  **Denkinger 2010** | **(4)**  **Gupta 2015** | **(5)**  **Lagersted-Olsen 2016** | **(6)**  **Shaw 2017** | **(7)**  **Smith 2013** |
| **Demographics** | | | | | | | |
| **Education** | No | **8-pt scale**: none-university  **+**  **3 category:** low, medium, high scale | Protocol plans to record. Not yet described how. | No | No | No formal qualifications/Basic e.g., O-levels, A-levels or equivalents/Advanced e.g., degree or professional qualification | Unclear |
| **Social Class** | No | No | No | No | No | No | No |
| **Income** | (< $30k/$30 to < $60k/$60k to < 100k/≥ $100k/refused or don’t know or missing) | No | No | No | No | No | No |
| **Ethnicity** | Aust, NZ, Indigenous, Other English speaking, Sth Europe, Other Europe, Asia, Other | No | No | No | No | No | No |
| **Employed/retired** | (full time/part time/retired/other not working/missing) | No | No | No | No | Yes | n/a |
| **Marital status** | Married or defacto | No | No | No | No | Married, Cohabiting, Single, Divorced, Separated, Widowed | No |
| **Area depirvation** | No | No | No | No | No | Carstairs measure of area deprivation based on the 2011 census; Scottish Index of Multiple Derivation (SIMD) 2012 | No |
| **Working conditions** | No | No | No | -job seniority: months, m(SD)  -influence at work: m (SD)  -occupational lifting/carrying time at work: n (%) | Seniority factor  Lift factor (self-reported kg and duration minutes per day) | No | Effort-Reward Imbalance Instrument |
| **Occupation** | *Present occupation or previous if not working* (managers or professionals/technical & trade or community & personal service/clerical & administrative or sales/machinery operator & driver or laborer/never worked or unknown ) | No | No | Yes | Yes | No | Yes |
| **Location** | Region/country | No | Region/country | Region/country | Region/country | Region/country | Region/country |
| **Setting** | Yes (community-based) | Yes (population-based) | Sheltered housing | Occupational | Occupational | Community and clinical practice | Occupational |
| **Social context** | No | No | LSNS-R; LSNS-6 | No | No | Social particiapnt measures (categorical lists) | Potential socioecological correlates questionnaire |
| **Environmental context** | No | No | ENABLE-AGE | No | No | Access domain, natural space, walkbility, pensioner density, population density.  Home environment: garden status, floors in house etc. | n/a |
| **Assests and Living arrangments** | Assests: ownership of current residence  Housing type | No | No | No | No | No | No |
| **Cardiometabolic outcomes** | | | | | | | |
| **BMI** | Yes | Yes | Yes | Yes | Yes | No | No |
| **Waist circumference** | Yes | Yes | Yes | No | No | No | Yes |
| **Glucose** | Yes (fasting and non-fasting) | Fasting (cut-off values of 126 mg/dl (7.0 mmol/L))  +  2-hours oral glucose tolerance test (cut-off values of 200 mg/dl (11.1mmol/L) | Yes (not clear if fasting) | No | No | No | No |
| **Insulin** | No | No | No | No | No | No | No |
| **Cholesterol, total** | Yes | No | No | No | No | No | No |
| **Cholesterol, HDL** | Yes | Yes | Yes | No | No | No | No |
| **Cholesterol, LDL** | Yes | Yes | Yes | No | No | No | No |
| **Triglycerides** | Yes | Yes | Yes | No | No | No | No |
| **Systolic Blood pressure** | Yes | Yes | Yes | No | No | No | Yes |
| **Diastolyic blood pressure** | Yes | Yes | Yes | No | No | No | Yes |
| **HbA1c** | Yes | No | No | No | No | No | No |
| **Blood cell count** | No | No | Yes | No | No | No | No |
| **Creatine** | No | No | Yes | No | No | No | No |
| **Gamma-glutamyltransferase** | No | No | Yes | No | No | No | No |
| **Uric acid** | No | No | Yes | No | No | No | No |
| **Urea** | No | No | Yes | No | No | No | No |
| **Albumin** | No | No | Yes | No | No | No | No |
| **Health Behaviours** | | | | | | | |
| **Self-reported Physical Activity** | No | No | SPPB  ADL  LASA PA Questionnaire  Diary | No | No | No | Diary/logbook  Route interviews  Movement at Work Questionnaire  EPAQ-2  Modifed Tecumesh Occupational Activity questionnaire |
| **Diet** | -Fiber intake (g/day)  Fat, %E  -Saturated fat, %E  -Sodium intake (mg/day)  -Potassium intake (mg/day)  -Fruit and vegetable serves (serves/day) | No | Mini Nutritional Assessment (MNA 18-item)  Dietary supplements | No | No | No | No |
| **Smoking** | Yes | never, former, current smoker | No | Yes | No | No | No |
| **Alcohol intake** | *Alcohol intake* (g/day) | -no use  -low alcohol use (women 7 glasses per week, men 14 glasses per week)  -high alcohol use (women >7 glasses per week, men >14 glasses per week) | Recall method of total intake: (weekday consumption x 5) + weekend consumption | No | No | No | No |
| **Health Status** | | | | | | | |
| **Physician-diagnosed disease** | -Prior CVD  -Family history of diabetes  -Mobility limitation | -Prevalence of T2DM  -Diabetic medication use (i.e. oral glucose lowering medication and insulin)  -History of CVD | Unclear from protocol | Unclear | Unclear/No | No | No |
| **Self-rated health** | SF-36 General Health and Well-Being questionnaire | No, only EuroQoL-5D | No | No | No | No | 12-item General Health Questionnaire (psychological health) |
| **Quality of life** | No | No | No | No | No | See social and environmental contexts | No |
| **Mobility** | Phyisician-diagnosed mobility limitation | -No mobility limitations  -Problems walking  EuroQoL-5D | No | Mobile Yes/No | No | No | No |
| **Pain** | No | No | ’structured pain interview’ | No | No | No | No |
| **Urinary and facecal incontinence** | No | No | ICIQ-SF | No | No | No | No |
| **Psychological distress, exhuastion** | No | No | 10-pt NRS | No | No | No | No |
| **Cognitition** | No | No | Category fluency, episodic fluency (CERAD battery), Letter sorting, cognitive estimation tasks | No | No | No | No |
| **Medication use** | *-*Contraceptive pill use (yes/no/not applicable [male])  -BP tablets (yes/no)  -Chol tablets (yes/no)  -Diabetes med (yes/no) | No | Barcode scan, medication calendar | No | No | No | No |
| **Falls** | No | No | SF of Falls Efficacy Scale International  Fatigue NRS 10-pt  Dizziness 5-pt Likert | No | No | No | No |
| **Low back pain intensity** | No | No | No | 10-pt NRS | Nordic Questionnaire for the Analysis of Musculoskeletal Symptoms | No | No |
| **Menopause** | post-menopausal/going through menopause/pre-menopausal/not applicable [male] | No | No | No | No | No | No |
| **Mental Health** | depression score category (scale unstated) | No | No | No | No | No | No |
| **BMI:** body mass index; **HDL:** high-density lipoproteins; **LDL:** low-density lipoproteins; **QoL:** quality of life; | | | | | | | |

# **Appendix 2: Workshop 2**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table 2: Some Identified Variables with Potential Harmonisation Issues | | | | |
| Variable Area | **Examples of measures taken** | | | |
| Demographics | | | | |
| Education level | Completed: school, apprenticeship/diploma, uni. degree | Completed: school, uni. degree, grad. uni. degree | Illiterate, read/write, 4-8yrs, 8-11yrs | Low, moderate, high |
| Marital status | Married | Married + de facto |  |  |
| Self-reported Physical Activity | | | | |
|  | Activity counts / METS | Categorised: sedentary, low active, active, very active | Duration (%) in posture: lying/sitting, standing, walking | Duration (%) in posture: lying, sitting, standing, dynamic standing, walking, running, cycling |
| Clinical and other Health Outcomes | | | | |
| General Health | Poor, fair, good, very good, excellent | 0-10 scale |  |  |
| Cardiovascular Health | Cholesterol | Blood glucose | HRV/HRR | Blood Pressure |
| Mental Health | Mini Mental State Examination (score <25) | DASS |  |  |
| Pain | 0-10 scale (worst, average, best) | Number of days with (bothersome) pain |  |  |
| Disability | EuroQol-5D | ÖMPQ | RMDQ | ODI |

**However:** Final decisions require a complete variable list.

# **Appendix 3: Workshop 3**

## Appendix 3.1: Summary of draft protocol for accelerometer measurement

**Prospective Harmonization of Accelerometer Data**

By taking advantage of the accelerometers’ capacity as an inclinometer and placing the accelerometer on the thigh, studies are now able to provide more detailed information on physical activity types (e.g. walking, cycling and running) and body postures (e.g. sitting/lying and standing) (Skotte, et al., 2014) (Stemland, et al., 2015). Large cohort studies have started to implement thigh mounted accelerometer protocols, primarily using three different brands of triaxial accelerometers: ActiGraph, ActivPAL and Axivity. To facilitate future pooling of individual participant accelerometer data, harmonization of methods for data collection is warranted (Atkin, et al., 2017) (Fortier, et al., 2015). The following key recommendations for accelerometer measurement protocols are based on a review by Edwardsson (Edwardson, et al., 2017), a best practice paper by Matthews et al (Matthews, et al., 2012), and on experiences from ongoing epidemiologic studies in free-living adults.

* **Wear period** Employ a 24-hour wearing protocol for at least 7 days: midnight to midnight.
* **Attachment location** Anterior mid right thigh.
* **Attachment method**
* Attachment methods: Hypafix, Opsite flexifix, Palstickie, ‘Hair-set’ double sided adhesive tape in combination with plastic covering and Opsite flexifix.
* If the accelerometer is not waterproof, it should be sealed in waterproof dressing/nitrile sleeve.
* It is not clear how the type of attachment may influence wear compliance.
* **Instructions for participants** Provide verbal, visual, and written instructions to participants on how to wear the device correctly.
* **Diary** Provide a diary (paper or electronic) to collect information on wake and sleep time, time in and out of bed, any removal times, and other contexts of interest, e.g. work times. Keep it simple.
* **Sample frequencies, software, initialization and data extraction** Default sample frequencies for the Actigraph, Axivity and ActivPal accelerometers.
* **Calibration** If multiple accelerometers are used simultaneously, a calibration procedure should be performed.
* **Compliance and return of monitors** How to distribute and have participants return monitor and diary after the field-based data collection period to avoid loss of monitors and valuable data.
* **Data quality control** Use quality controls to check classifications (wake time, work time, sleep etc), ideally against an external source of data, such as a diary. No waking wear identification method is universally accepted.

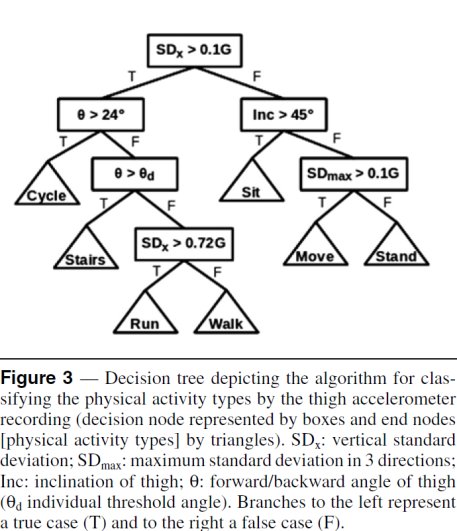
## Appendix 3.2: Tables of core variables and core outcomes

|  |  |  |  |
| --- | --- | --- | --- |
| Table 3: Core Contextual Variables | | | |
| Domain | **Construct** | **Description** | **Challenges/comments** |
| Sociodemographic | Gender |  |  |
| Age |  |  |
| Ethnicity |  | Country specific |
| Marital status | Married/co-habituating |  |
| Work | Work (yes/no)  Type of occupation  Occupation class | Country- and cultural specific |
| SES | Education | Level/years of education | Country- and cultural specific |
| Income | Self-report/register | Country specific |
| Behaviour | Physical activity | Self-reported sports/exercise  Self-reported commuting |  |
| Sedentary behaviour | TV-viewing time  Pc-use  Screen time |  |
| Sleep | Duration/quality |  |
| Smoking | Daily/occational/never smoker/ex- smoker |  |
| Alcohol | Units per week consumption |  |
| Diet | Dietary quality |  |

|  |  |  |
| --- | --- | --- |
| Table 4: Core intermediate health outcomes | | |
| Domain | **Construct** | **Description** |
| Anthropometrics | Weight |  |
| Height |  |
| BMI |  |
| Waist circumference |  |
| Fat percentage |  |
| Cardiometabolic biomarkers | Blood lipids | Total, LDL,HDL, Trigyceride |
| Glucose | Plasma/Blood |
| Blood pressure |  |
| Health-related | Functional limitation | Self-report |
| Self-reported health |  |

# **Appendix 4: Workshop 4**

## Appendix 4.1: Acti4 Software



**Figure 1:** Decision tree used in Acti4 software. Used with permission (Skotte, et al., 2014)

## Appendix 4.2: Pros & cons of data processing options

|  |  |
| --- | --- |
| Table 5: Commercial software | |
| Pros | **Cons** |
| Ease of use | Proprietary algorithms (Often black box) |
| Rapid result processing | Fixed/Limited classification ability |
| Uniform outcome measures | Misclassification during daily living |
| Uniform file formats | Less flexible to data handling |
| New software versions developed | Typically product license is required |
| - | Limited to the commercial accelerometer |
| “Who? What? When? Where?”: Data can be processed remotely as all outcome variables will be the same, provided recording settings are the same. Data files do not require format modification. Low time resource requirements, moderate financial resources required (if software is purchased) | |

|  |  |
| --- | --- |
| Table 6: Rule-based software | |
| Pros | **Cons** |
| Broad activity classification | Some training required |
| Low misclassification during daily living | Currently considerable manual work (diary) |
| Adaptable to multiple input file formats and accelerometer manufacturers | File conversion is time consuming |
| Transparent processing process | Reduced feasibility as sample size grows |
| The Acti4 software is free to share for ProPASS | - |
| “Who? What? When? Where?”: Data can be processed remotely, provided software training is given. Data files require format modification before processing. Moderate time resources (for processing) and moderate financial resources required (for software development and training) | |

|  |  |
| --- | --- |
| Table 7: Machine-learning | |
| Pros | **Cons** |
| Potential for accurate classification | Requires highly specific skill set/knowledge and resources to develop new approach |
| Broad activity classification | Requires a considerable training set to develop or generalize to other accelerometers/populations |
| Reduced manual work | If developed, large time/resources needed to develop for different accelerometer manufacturers and potentially populations |
| Capable of handling larger sample sizes | - |
| “Who? What? When? Where?”: Data must be processed centrally. Data files may require format modification before processing. Low time resources (automated analysis) and high financial resources (development of the method and development of the software. | |

## Appendix 4.3: Summary of data processing methods in reviewed studies

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Table 8: A *summary the data processing methods used in the reviewed studies.* | | | | | | |
| Authors | **Main objective(s)** | **Programming software** | **Downloading software** | **Data cleaning methods** | **Data processing software** | **Extracted variable groups** |
| (Aguilar-Farias, et al., 2015) | To compare the reliability and validity of two self-report instruments to ActivPAL measurements | ActivPal a | ActivPal | A semi-automated filter ensuring only waking hours and wear time are included | ActivPal | Time spent in sitting/lying postures |
| (Aguilar-Farias, et al. 2014) | To determine cut-points with the highest accuracy for estimating time spent in SB | ActivPal b | ActivPal | Actigraph data was downloaded using the low filter extension. | ActivPal | ST (sitting/lying, standing, upright but not walking) |
| (Barreira, et al., 2015) | To compare free living accelerometer derived and posture derived estimates of breaks in ST. | not specified | not specified | Data was processed using the low frequency extension filter integrated into 60 second epochs. Actigraph breaks were computed when a transition occurred from a minute with <100 activity counts to an adjacent following minute with >100 activity counts | ActiLife, ActivPAL | - |
| (Clark, et al., 2018) | To assess Bluetooth derived proximity data accuracy of location and to assess the effect of sensor positioning | not specified | not specified | not specified | not specified | Total time spent sitting, standing, stepping, purposeful walking |
| (Clemes, et al., 2012) | To determine the concurrent validity of inclinometer and different CPM cut points for detecting SB | not specified | not specified | The output from accelerometers was re-integrated into 60 second epochs | not specified | ST based on CPM |
| (Cooper, et al., 2017) | To determine the accuracy of wrist mounted accelerometer vs thigh mounted sensors | not specified | not specified | A three step process was applied to train the classifier: windowing (partitioning the data stream into time lengths), labelling (each window by activity), feature extraction (producing characteristic descriptors from the raw accelerometer signal) | not specified | Characteristic PA descriptors |
| (Denkinger, 2012) | To determine the mediating effect of physical activity on the association of centrally acting medication with falls | not specified | not specified | not specified | not specified | Total walking time, mean number of falls |
| (Deshmukh, et al., 2012) | To compare COP-based measures of postural stability to those acquired by inertial sensors | not specified | not specified | A fourth order low pass Butterworth filter with a 10Hz cut-off frequency | not specified | Postural sway |
| (Florencio, et al., 2015) | To explore physical activity patterns of women living in a poor socio-economic environment | not specified | not specified | not specified |  | SB, low active, active, high active |
| (Fortune, et al., 2015) | To determine and compare the validity of using multiple activity monitors versus a single monitor to count steps | not specified | not specified | Data was filtered and upright dynamic activity was calculated using both angles and SMA. Posture was detected using waist and thigh data. | not specified | Walking/fidgeting, jogging, number of steps |
| (Gao, et al., 2017) | To assess short term and long term recall against thigh-mounted accelerometry | not specified | not specified | Data was transformed into a polar coordinate system. Inclination in the sagittal plane was filtered at a cut-off of 1Hz, sitting and upright positions were classified as a thigh inclination angle Sedentary and upright postures were divided based on a thigh angle of 45 degrees from the horizontal and vice versa. Minimum detection window for the classification of postures was 5 seconds. | OpenSALTO | Sitting, standing, walking |
| (Gupta, et al., 2016a) | To investigate objectively measured temporal patterns of sitting | not specified | not specified | not specified | Acti4 | MVPA , SB bouts long, moderate, brief |
| (Gupta, et al., 2016b) | To investigate the association of obesity indicators with total sedentary time and bouts of sedentary time | not specified | not specified | Data was low pass filtered with a 5Hz 4th order Butterworth, and split into 2 second intervals with 50% overlap. A sedentary posture is that in which the inclination of the thigh is above 45 degrees. The daily reference measurements (i.e. standing in an upright position for 15 sec on every measure day) were used to obtain the angle between the leg and accelerometer axis (i.e. accelerometer position) | Acti4 | Sitting, lying, standing ,MVPA, SB long, moderate, brief |
| (Gupta, et al., 2018) | To investigate if age, gender, BMI and LBP introduce differential bias in self-reported information on ST and PA | not specified | not specified | not specified | Acti4 | LPA, MVPA |
| (Hallman et al. 2017) | To investigate if OPA and LPTA are differentially associated with cardiac autonomic modulation during sleep | ActiLife b | not specified | not specified | Acti4 | OPA,LPA |
| (Hallmann, et al., 2015a) | To determine the association between NSP intensity and sitting time per day | ActiLife b | ActiLife | Data was low-pass filtered with a 5-Hz fourth-order Butterworth filter and then split up into 2-s sequences with 50 % overlap | Acti4 | LST, OST low, medium, high |
| Hallman et al. 2015b) | To document temporal patterns of OPA and LPA | ActiLife b | ActiLife | Low-pass filtered using a 5 Hz 4th order Butterworth filter and then split up into 2 s sequences with 50 % overlap | Acti4 | Bouts of sitting, standing, walking, running, cycling |
| (Hallman, et al., 2016) | To investigate temporal patterns of OPA and LPA sitting and the association with neck-shoulder pain | ActiLife b | ActiLife | Low-pass filtered using a 5 Hz 4th order Butterworth filter and then split up into 2 s sequences with 50 % overlap | Acti4 | SB bursts, moderate, prolonged |
| (Hallman, et al., 2015c) | To assess sitting time across several working days while resting HRV is assessed during sleep | ActiLife b | ActiLife | not specified | Acti4 | Time spent sitting, lying, walking, walking fast pace, running, cycling, walking stairs |
| (Klenk, et al., 2013) | To analyze the association of daily walking duration with cardiovascular biomarkers of inflammation | not specified | not specified | not specified | not specified | Time spent lying/sitting, standing, walking |
| (Klenk, et al., 2015) | To analyze the association between objectively measured PA and vitamin D status | not specified | not specified | not specified | not specified | Time spent lying/sitting, standing, walking |
| (Kloster, et al., 2017) | To address different definitions of minimum break length on TST, sit-to-stand transitions, number of prolonged sitting periods, prolonged sitting periods | not specified | not specified | not specified | Acti4 | Sit-to-stand transitions, bouts of sitting time and activity |
| (Koster, et al., 2016) | To compare the ST estimates from ActivPAL and hip-/wrist worn ActiGraph | not specified | not specified | not specified | not specified | SB |
| (Kurita, et al., 2017) | To compare the waist-worn Active style Pro HJA-350IT, the waist-worn ActiGragh™GT3X+ and the thigh-worn activPAL3 in assessing SB | not specified | not specified | Time-stamped 'event' data was used to calculate sedentary bouts were determined by any code of sitting/reclining | PAL Analysis | TST, SB bout duration, number of breaks in SB |
| (Liu & Chang, 2009) | To consider the optimal position for the accelerometer, recognition of; walking, sitting down, and falling, an identification of a device suitable for use in medical applications | LabVIEW software | not specified | Raw accelerometer signals were converted into polar coordinates and a the feature vector was extracted | Self-constructing neural fuzzy inference network (SONFIN) | Time spent sitting, walking, number of falls |
| (Lyons, et al., 2005) | To describe an accelerometer based mobility monitoring technique to distinguish between static and dynamic activities | not specified | not specified | not specified | Matlab | Time spent sitting, standing, lying, moving |
| (Mannini & Sabatini, 2011) | To classify sit, stand, cycle walk and run using support vector machines | ActiNav | ActiNav | Data were windowed (250 points included within each window, with 50% overlap) and feature vectors were evaluated for each window: mean median, variance, peak, and ranges. Pearson’s correlation between axes, stride time, and biometric data, as well as coefficients of Fast Fourier Transforms was computed. Two cascade SVM classifiers were used to obtain the activity label and the point estimate of the locomotion speed for feature vectors classified with walk or run labels. | LibSVM package, MatLab | Time spent sitting standing walking running, cycling |
| (Pulakka, et al., 2018) | To examine the association between employment status and SB and PA | ActivPAL c | ActivPAL | Normally distributed descriptive variables were summarized as means with standard deviations, variables with skewed distribution as medians and interquartile ranges, and categorical variables as numbers and percentages. | Matlab R2013b | TST, time spent walking, standing, stepping |
| (Riou, et al., 2015) | To build a classification model for biaxial and tri-axial accelerometers and validate the performance of this classification model | not specified | INTERVIEW | A training set was built using 22 participants and activity classifications were based on recording time. Transitions were removed from training sets. Two classification models, using recorded features and known activity, were built using this training sample for biaxial and tri-axial accelerometers respectively. | Activity Recognition | Time spent lying down, dynamic standing, sitting, walking and running |
| (Schneller, et al., 2015) | To compare the accuracy of EE estimation using new and widely used methods against indirect calorimetry | ActivPAL d | ActivPAL | METS, CPM , SB, and posture were calculated using respective software algorithms | ActivPAL, ActiLife, Acti4 | Lying/sitting, standing, stepping, walking, running, cycling, walking stairs, move, CPM |
| (Skarpsno, et al., 2018) | To investigate the associations between OPA and LTPA and insomnia/non-restorative sleep. | ActiLife b | ActiLife | not specified | Acti4 | OPA and LTPA (high, medium, low) |
| (Skotte, et al., 2014) | To develop a method for identifying everyday PA types using 3D accelerometers | ActiLife b | ActiLife | Low pass 4th order Butterworth; 5Hz. Median filtering was carried out post-hoc. Data was averaged over 2 seconds, achieved by taking a predetermined number of cycles per movement. Mean acceleration was calculated as was standard deviation of acceleration. Inclination of the X-axis was calculated, and assumed parallel to the axis of the thigh, therefore the inclination gives the angle between the vertical line and thigh axis (0-180) | Acti4 | Time spent walking, running, sitting standing, cycling, stair walking |
| (Steeves, et al., 2015) | To determine whether the ActiGraph and ActivPAL monitors could successfully identify PA type compared to direct observation. | ActiLife e, ActivPAL Research Edition f | not specified | Data was processed a 1s epoch level, and event data from ActivPAL was converted to second by second data | ActiLife, ActivPAL | Time spent sitting, standing, walking, cycling |
| (Wullems, et al., 2017) | To examine the overall balanced accuracy and robustness of four heterogeneous pooled-data algorithms | not specified | not specified | Signal was filtered per axis twice using a zero-phase 4th order low pass Butterworth filter at 20hz and 0.5 Hz. The following were calculated over 10s windows: arithmetic mean, standard deviation (SD), minimum, maximum, median, interquartile range (IQR), skewness, kurtosis, root mean square, cross-correlation, roll, pitch, yaw, peak-to-peak amplitude, peak intensity, zero- crossings, lag one autocorrelation, dominant frequency, amplitude of dominant frequency and entropy. 10 second windows were further used to develop three cut-off point based algorithms using: the sum of vector magnitudes, the summation of time integrals of the moduli of tri-axial signals, based on postural balance using total movement calculations. The final algorithm used an ensemble Random Forest method. | R 3.2.5 | Time spent lying, sitting, standing, shuffling sideways, walking, cycling, and treadmill walking, EE |
| *(ST = sedentary time, SB = sedentary behaviour, PA = physical activity, CPM = counts-per-minute, TST= total sitting time, OPA = occupation physical activity, LTPA = leisure time physical activity, OST = occupational sedentary time, LST =leisure sedentary time, EE = energy expenditure, MET = metabolic equivalents, LPA = light physical activity, MVPA = moderate to vigorous physical activity)* | | | | | | |
| 1: v6.1.2 Research Edition Pal Technologies Ltd., 2010, 2: v5.5 Actigraph LLC, Pensacola, FL, USA 3: version not specified, 4: v7.1.18, PAL Technologies Ltd., Glasgow, Scotland; ActiLife v6.5.2, ActiGraph, TM, Pensacola, FL, USA, 5: v 6.5.3, 6: version 6.4.1, 7: v6.0 ActiGraph LLC., Pensacola, Florida, USA, ActivPAL3 v6.4 PAL Technologies Ltd, Glasgow, Scotland, 8: v. 7.2.32 | | | | | | |

# **Appendix 5: Workshop 5**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 9: Data linkage in ProPASS | | | | | | | | |
|  | **Australian Longitudinal Study on Women’s Health (ALSWH)** | **1970 British Birth Cohort Study** | **Copenhagen City Heart Study** | **Finnish Retirement and Aging Study (FIREA)** | **Health2016** | **The Maastricht Study** | **The Nord-Trøndelag Health Study (HUNT)** | **SCAPIS Uppsala** |
| Data linkage sources | |  |  |  |  |  |  |  |
| Registry A: | Admitted Patients/Hospital data | NHS mortality record | Mortality/cause specific | Pop Register primary health care visits | - | Hospital records | Hospital records | - |
| Registry B: | MBS/PBS | Hospital episodes (HESS) | Hospitalisation | Mortality: Statistics Finland, Finnish Envir. Institute | - | Mortality: Netherlands Bureau of Statistics | Cause of Death Registry | - |
| Registry C: | Cancer, Perinatal | - | - | Registry of Keva, Pension Institute | - | - | Cancer registry | - |
| Once your study data are linked to administrative health and mortality records, can they be exported to another country in anonymised individual level format? | | | | | | | | |
| Yes/No | Yes | Yes | No | No | Yes | Yes | Yes | No |
| Details | Must be onsite (Australia) or offsite via remote access (SURE). | Involves special licence. All cases reviewed on individual basis. | - | - | - | Likely | EU countries under GDPR, and non-EU with additional clauses in DTA | - |
| Current laws/policies regarding data sharing in your state/country, or sharing culture of your institution. Can the data be accessed for: | | | | | | | | |
| Export to countries & be pooled |  | Don’t know |  |  |  |  |  | - |
| No export but remotely access as IPD | - | - | - |  | - | - |  | - |
| For federated analyses\* | - | - |  | - | - | - |  | - |
| No remote access to IPD | - | - | - | - | - | - | - | - |
| No, only group-level estimates \*\* | - | - | - | - |  | - | - | - |
| Other | Only accessed remotely via SURE | - | - | - | Possible, but currently not an option | - | Provide files (delete when complete) | No sharing is possible at the moment |
| * : selected; (-): not selected or no information provided; IPD: individual patient level data; | | | | | | | | |

\*The data can be accessed for federated analyses, i.e. script-based analyses are done in your servers and individual level estimates are combined with other such estimates

\*\*No data exportation or access from remote analysts is allowed, you can only provide group-level estimates from your database on a protocol that the consortium will provide**.**

# **Appendix 6: Workshop 6**

## Appendix 6.1: Figures of knowledge hierarchy

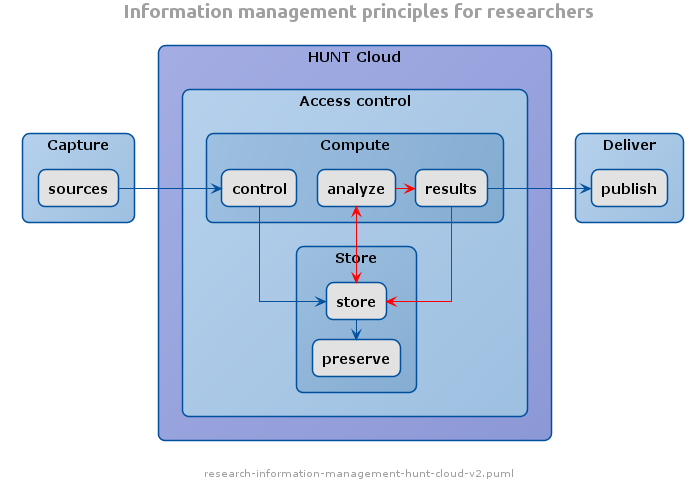
**Figure 2:** Knowledge Hierarchy - There is a great need for in-depth expertise to handle the changing and needs in data management - Knowledge management; one process of transforming data in wisdom (“mind the gaps” in the knowledge hierarchy) <https://www.quora.com/Can-someone-give-a-simple-example-for-data-information-knowledge-and-wisdom>

****

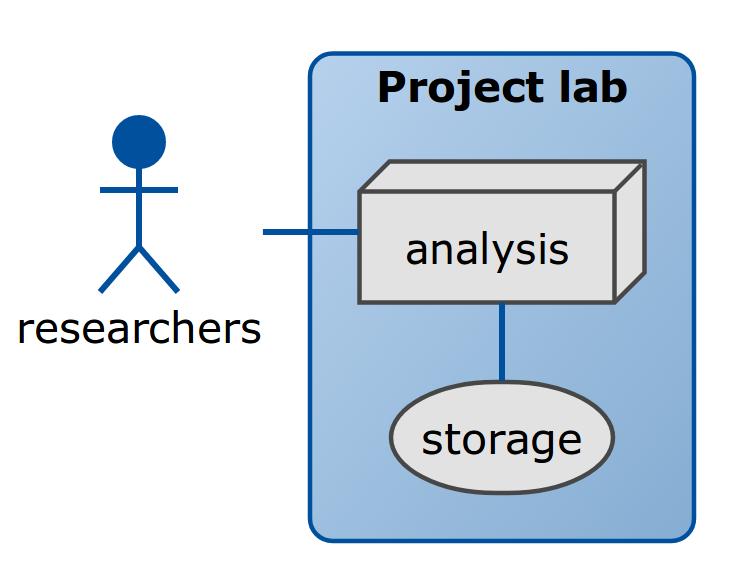
**Figure 3:** Rule of thumb for knowledge intensive organizations are to invest 80% effort into people and process.

## Appendix 6.2: Data sharing platforms

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table 10: Types of Data sharing platforms | | | | |
|  | **Centralized** | **Federated** | | **Decentralized:** |
|  |  | Semi-distributed/separated | Fully distributed |  |
| Data access | Access to all data/studies in unified manner | Access to multiple studies through central portal.  "Access to studies according to their individual sharing policy" | Access to each study separately, but under a common usage/access policy and single approval. | Limited access, ad hoc “coordination” |
| Data analytic | Most powerful search, analysis, quality assurance of aggregated data | Cross-repository searching and analytics; Metadata and aggregate statistics can be developed by central authority | Index/Catalog only | Index/Catalog only |
| Deployment costs | High.  Structure and build centralized storage; Develop data interoperability mechanisms; Develop common usage policy and governance system | Moderate.  Develop data interoperability mechanisms; Develop common usage policy and governance system | Moderate.  Develop common usage policy and governance system | Low |
| Operational costs | Moderate.  Operating and maintaining central storage; administering policies.  Few distributions cost | Moderate  Operating and maintaining portal; administering policies | Moderate  Operating and administering policies. | Low |



**Figure 4:** HUNT cloud management, from sources to publish. HUNT Cloud (Norwegian University of Science and Technology, NTNU).



**Figure 5:** HUNT cloud digital labs.

# **Appendix 7: Survey of Main Cohorts**

**Date of survey:** September 2018

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 11: Characteristics of the main cohorts | | | | | | | | |
|  | **Australian Longitudinal Study on Women’s Health** | **1970 British Birth Cohort Study** | **Copenhagen City Heart Study** | **Finnish Retirement and Aging Study (FIREA)** | **Health2016** | **The Maastricht Study** | **The Nord-Trøndelag Health Study (HUNT)** | **SCAPIS Uppsala** |
| Principal investigator | Professor Gita Mishra | M Hamer and A Sullivan | Andreas Holtermann | Sari Stenholm | Allan Linneberg | Annemarie Koster | Paul Jarle Mork | Johan Sundström |
| Accelerometry Contact Person | Manos Stamatakis | Mark Hamer | Andreas Holtermann | Anna Pulakka | Mette Aadahl | Coen Stehouwer | Steinar Krokstad | Peter Palm |
| Location of leading institution | The University of Queensland | Loughborough University & UCL | National Research Centre for the Working Environment, Copenhagen | Turku, Finland | Copenhagen | Maastricht University, the Netherlands | Norwegian University of Science and Technology | Uppsala University |
| Geographical coverage of the study | Australia | Great Britain | Two districts of Copenhagen | Southwest Finland | Western part of greater Copenhagen | South of the Netherlands | Northern part of Trøndelag region | Uppsala Region |
| Accelerometry number (estim.) | 1,500 | 6,500 | 2,032 | 280 | 1,251 | 8,000 | 50,000 | 5,000 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 12: Data Access | | | | | | | | |
|  | **Australian Longitudinal Study on Women’s Health** | **1970 British Birth Cohort Study** | **Copenhagen City Heart Study** | **Finnish Retirement and Aging Study (FIREA)** | **Health2016** | **The Maastricht Study** | **The Nord-Trøndelag Health Study (HUNT)** | **SCAPIS Uppsala** |
| Which individuals or bodies need to approve data sharing for: | | | | | | | | |
| *Accel. data* | * PI * ALSWH Data Access Committee | * PI & Co-PI * Institution | * PI & Co-PI * Institution * Other **\*** | * PI & Co-PI | * PI & Co-PI * Institution * Mette Aadahl | * PI & Co-PI | * PI & Co-PI * Institution * Ethics * Other **\*\*** | * PI & Co-PI * Ethics * Steering committee |
| *Non-accel. data* | * ALSWH Data Access Committee | * None - open access | * PI & Co-PI * Institution * Other \* | * PI & Co-PI | * PI & Co-PI * Institution | * PI & Co-PI | * PI & Co-PI * Institution * Ethics | * PI & Co-PI * Ethics * Steering committee |
| Likelihood that required permissions will be secured to share: | | | | | | | | |
| *Accel. data* | Almost certain | Very likely | Very likely | Almost certain | Very likely | Almost certain | Unknown | Not sure |
| *Non-accel. data* | Very Likely | Almost certain | Likely | Almost certain | Very likely | Very Likely | Unknown | Rather unlikely |
| Export data (anonymised individual level format) | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Not sure |
| Additional permissions required | No | No | No | No | No | Yes (sharing contracts to be signed) | GDPR | Not sure |
| Timing of data availability | 2021-2022 | 2019 | 2019 | Now | 2021-2022 | Now (n=3451)  Later (N=7000) | Autumn 2019 | 2021? |
| (n): number; PI: Principal Investigator; | | | | | | | | |

**\*Other (Cop City):** scientific committee of Copenhagen city heart study and NRCWE

**\*\*Other (HUNT):** Principal investigator has exclusive rights for a defined period of time, after which all data will become part of the larger study.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 13: Data Linkage | | | | | | | | |
|  | **Australian Longitudinal Study on Women’s Health** | **1970 British Birth Cohort Study** | **Copenhagen City Heart Study** | **Finnish Retirement and Aging Study (FIREA)** | **Health2016** | **The Maastricht Study** | **The Nord-Trøndelag Health Study (HUNT)** | **SCAPIS Uppsala** |
| Data linkage sources | | | | | | | | |
| Registry A: | Admitted Patients/ Hospital data | NHS mortality, Hospital episodes (HESS) | Mortality | Pop Register of primary health care visits | - | Hospital records | Hospital records | - |
| Registry B: | MBS/PBS | - | Cause-specific mortality | Statistics Finland, Finnish Environmental Inst | - | Netherlands Bureau of Statistics | Cause of Death Registry | - |
| Registry C: | Cancer, Perinatal | - | Hospitalisation | Registry of Keva, Pension Institute | - | - | Cancer registry | - |
| Registry D: | National Death Index, Cause of Death | - | - | - | - | - | - | - |
| Once your study data are linked to administrative health and mortality records, can they be exported to another country in anonymised individual level format? | | | | | | | | |
| Yes/No | Yes | Yes | No | No | Yes | Yes | Yes | No |
| Details | Must be onsite (Australia) or offsite via remote access (SURE). | Involves special licence. All cases reviewed on individual basis. | - | - | - | Likely | EU countries under GDPR, and non-EU with additional clauses in DTA | - |
| Current laws/policies regarding data sharing in your state/country, or sharing culture of your institution. Can the data be accessed for: | | | | | | | | |
| Export to countries & be pooled |  | Don’t know |  |  |  |  |  | - |
| No export but remotely access as IPD | - | - | - |  | - | - |  | - |
| For federated analyses\* | - | - |  | - | - | - |  | - |
| No remote access to IPD | - | - | - | - | - | - | - | - |
| No, only group-level estimates \*\* | - | - | - | - |  | - | - | - |
| Other | Only accessed remotely via SURE | - | - | - | Possible, but currently not an option | - | Provide files (delete when complete) | No sharing is possible at the moment |
| * : selected; (-): not selected or no information provided; IPD: individual patient level data | | | | | | | | |

\*Data can be accessed for federated analyses, i.e. script-based analyses done in your servers and individual level estimates are combined with other estimates

\*\*No data exportation or access from remote analysts allowed, can only provide group-level estimates from your database (protocol provided by consortium)