Data harmonisation the International Children's Accelerometry Database (ICAD): Challenges and Examples

MRC Epidemiology Unit

Loughborough University

Ulf Ekelund (@Ulf_Ekelund)
ulf.ekelund@nih.no

NIH Department of Sports Medicine
NORWEGIAN SCHOOL OF SPORT SCIENCES

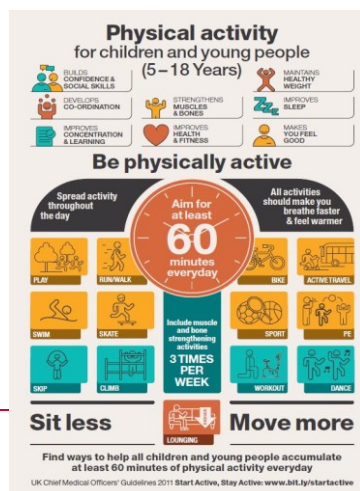
NIH Department of Sports Medicine
NORWEGIAN SCHOOL OF SPORT SCIENCES

Outline

- The International Children's Accelerometry Database (ICAD)
  - Background (ICAD 1.0)
  - Expansion (ICAD 2.0)
- Data harmonisation
  - Phenotypic data
  - Accelerometer data
  - Data availability

Why a pooled data base on childrens physical activity? ICAD 1.0

- **What we knew**



- **What we didn`t know**
  - Population levels and cultural differences
  - Dose-response:
    - Frequency, Intensity, Duration, Mode
  - Inter-relations between movement behaviours, sedentary behaviour and sleep)
  - Effect modifiers: population subgroups
  - Correlates and Determinants

NIH Department of Sports Medicine NORWEGIAN SCHOOL OF SPORT SCIENCES

Background for ICAD 1.0

- **Data availability** – Accelerometer data collected in many studies
- **Variability in data cleaning and reduction methods precludes comparison** (e.g. Cain et al. 2013)
- **ICAD** aimed to harmonise raw Actigraph accelerometer data, socio-demographic and phenotypic data using standardised methods to create comparable exposure and outcome variables across studies to improve our knowledge regarding the **strength of relationships between physical activity and health and to better understand the correlates and determinants of movement behaviors**.

Background

- **Individual person data pooling**
  - Increase statistical power – mediation / interaction analyses
  - create a more heterogeneous and potentially more representative sample
  - standardize and optimize the analytical methods used in the generation of outcome variables
  - provide a means to study the causes of inter-study variability in physical activity

NIH Department of Sports Medicine NORWEGIAN SCHOOL OF SPORT SCIENCES

Background

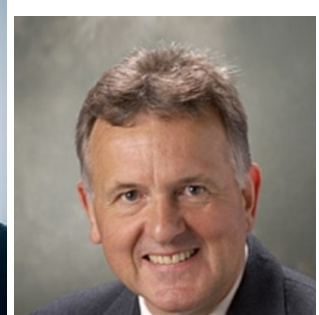
- ICAD established in 2007 with funding from the National Prevention Research Initiative (NPRI)

- Collaboration between University of Bath, University of Bristol and the MRC Epidemiology Unit, Cambridge

  - Chris Riddoch
  - Ulf Ekelund
  - Ken Judge
  - Ashley Cooper

The Original ICAD Team






NIH Department of Sports Medicine
NORWEGIAN SCHOOL OF SPORT SCIENCES

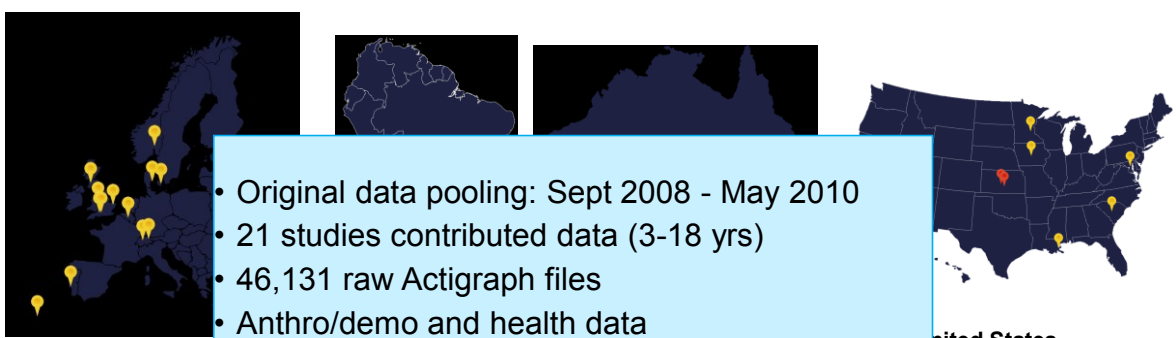
Sherar *et al. BMC Public Health* 2011, **11**:485
http://www.biomedcentral.com/1471-2458/11/485

BMC Public Health

RESEARCH ARTICLE Open Access

International children's accelerometry database (ICAD): Design and methods

Lauren B Sherar[1*], Pippa Griew[2], Dale W Esliger[1], Ashley R Cooper[3], Ulf Ekelund[4,5], Ken Judge[6] and Chris Riddoch[6]



- Original data pooling: Sept 2008 - May 2010
- 21 studies contributed data (3-18 yrs)
- 46,131 raw Actigraph files
- Anthro/demo and health data
- 10 XC, 7 Cohort, 4 Intervention

(Sherar et al, 2011)

13 European
- ALSPAC (U
- Ballabeina (
- Belgium Pre
- CHAMPS (UK)
- CoSCIS (DEN)
- EYHS (Den, Est, Por, Nor)
- KISS (CHE)
- MAGIC (UK)
- PEACH (UK)
- SPEEDY (UK)

nited States
DS
HANES 03/04
HANES 05/05

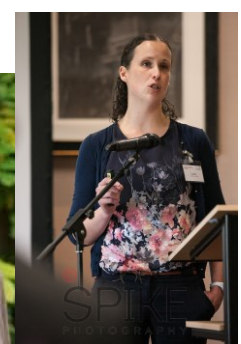
NIH Department of Sports Medicine
NORWEGIAN SCHOOL OF SPORT SCIENCES

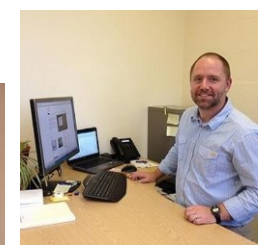
Current expansion ICAD 2.0

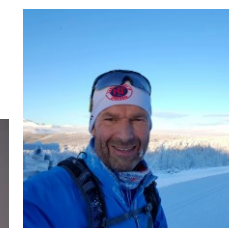
- Aims
  - Add **additional waves of accelerometer data and a wider range of non-accelerometer data** from existing studies
  - Harmonize a broader range of phenotypic information from existing studies http://www.mrc-epid.cam.ac.uk/research/studies/icad/data-harmonisation/
- Strategy
  - New waves of data submitted: 13 studies with at least two time-points
  - Additional variables from original submission: 7 studies
  - Approx. 55.000 accelerometer data files processed
  - 13 studies

ICAD 2.0 Working group









NIH Department of Sports Medicine NORWEGIAN SCHOOL OF SPORT SCIENCES

Data harmonisation

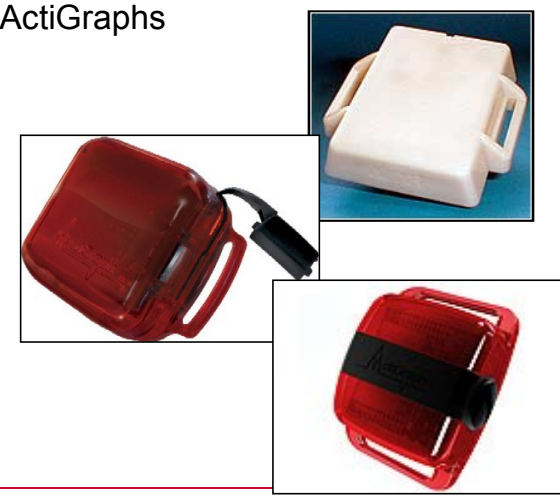
- Aim to attain, or at least improve, the comparability of information collected from different sources
  - Get your data ducks swimming in the same direction
- Non-accelerometer data
  - Retrospective data harmonisation requires clear, detailed notes on all variables in each study
  - >11,000 variables across 30 different constructs
- Currently Harmonised variables
  - Anthropometric, demographic, health, and correlates/determinants

Data harmonisation

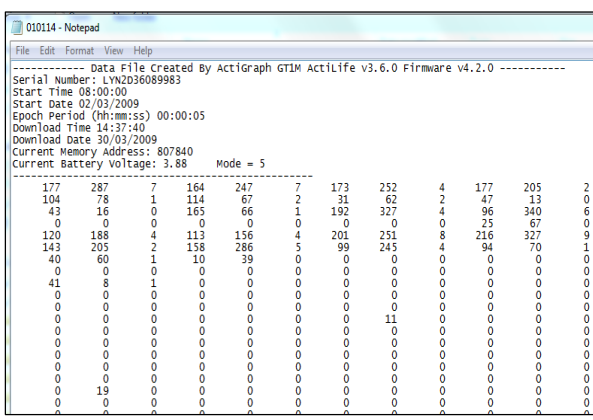
- Accelerometer data
  - ~55.000 accelerometer files from 21 studies
  - 49 waves of data collection from these studies
    - 49+ variations of initialization and deployment strategies
  - **Three step process**

STEP 1 – Standardising input data

- All file formats from any generation of ActiGraphs
  - CSA, MTI, GT1M, GT3X, GT3X+
  - Analysed by the Kinesoft software



STEP 1 – Standardising input data (60 sec epoch and vertical axis)



010114 - Notepad

File Edit Format View Help

```
---------- Data File Created By ActiGraph GT1M ActiLife v3.6.0 Firmware v4.2.0 ----------
Serial Number: LYN2D36089983
Start Time 08:00:00
Start Date 02/03/2009
Epoch Period (hh:mm:ss) 00:00:05
Download Time 14:37:40
Download Date 30/03/2009
Current Memory Address: 807840
Current Battery Voltage: 3.88     Mode = 5
-----------------------------------------------------------
```

177	287	7	164	247	7	173	252	4	177	205	2
104	78	1	114	67	2	31	62	2	47	13	0
43	16	0	165	66	1	192	327	4	96	340	6
0	0	0	0	0	0	0	0	0	25	67	0
120	188	4	113	156	4	201	251	8	216	327	9
143	205	2	158	286	5	99	245	4	94	70	1
40	60	1	10	39	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
41	8	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	11	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	19	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0

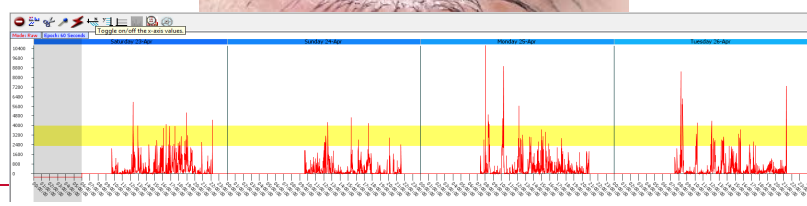
Example:
DAT-file (GT1M) with header mode 5: initialised in mode 5 (Vertical axis, 2nd axis, steps – three streams of data in the file)

- Original DAT exported to an AGD
- The AGD file exported to CSV
- CSV file opened in Excel
- Deleted all but the first column of data in the CSV file
- Changed the mode from 5 to 0 in the ninth line of the CSV file
- Saved the CSV file
- Converted the CSV file to AGD in ActiLife
- Exported the new AGD to DAT in ActiLife
- Reintgrated the new DAT file to 60 second epoch in Kinesoft

STEP 2 – Establish the validity

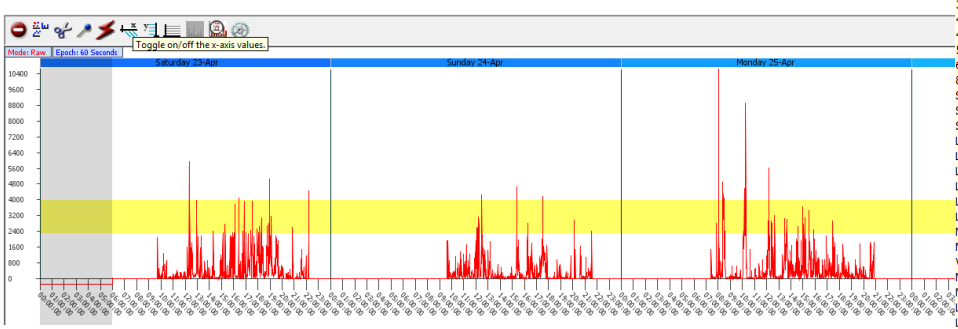
- **Each single accelerometer file manually examined for validity**

Visual inspection



STEP 2 – Establish the validity

- **Valid file (coded 0)**



Min Value:	0
Max Value:	10641
Average Value:	186
Time Spent in Cutpoints:	

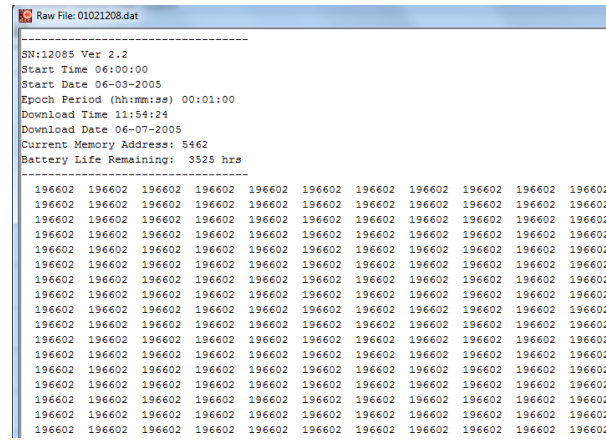
	Epochs	Minutes
0_50:	3567	3567.0
50_100:	330	330.0
100_150:	229	229.0
150_250:	282	282.0
250_500:	302	302.0
500_750:	196	196.0
750_1000:	140	140.0
1000_1500:	149	149.0
1500_2000:	100	100.0
2000_2500:	68	68.0
2500_3000:	46	46.0
3000_3500:	15	15.0
3500_4000:	13	13.0
4000_4500:	10	10.0
4500_5000:	4	4.0
5000_6000:	6	6.0
6000_8000:	2	2.0
8000_up:	3	3.0
Sedentary_50:	3567	3567.0
Sedentary_100:	3897	3897.0
Sedentary_150:	4126	4126.0
LPA_50_2000:	1728	1728.0
LPA_100_2000:	1398	1398.0
LPA_150_2000:	1169	1169.0
LPA_50_3000:	1842	1842.0
LPA_100_3000:	1512	1512.0
LPA_150_3000:	1283	1283.0
MPA_2000_6000:	162	162.0
MPA_3000_6000:	48	48.0
VPA_6000_up:	5	5.0
MVPA_2000_up:	167	167.0
MVPA_3000_up:	53	53.0
LVPA_50:	1895	1895.0
LVPA_100:	1565	1565.0
LVPA_150:	1336	1336.0
Evenson_SED:	3903	3903.0
Evenson_LPA:	1435	1435.0
Evenson_MPA:	99	99.0
Evenson_VPA:	25	25.0
Evenson_MVPA:	124	124.0

NIH Department of Sports Medicine
NORWEGIAN SCHOOL OF SPORT SCIENCES

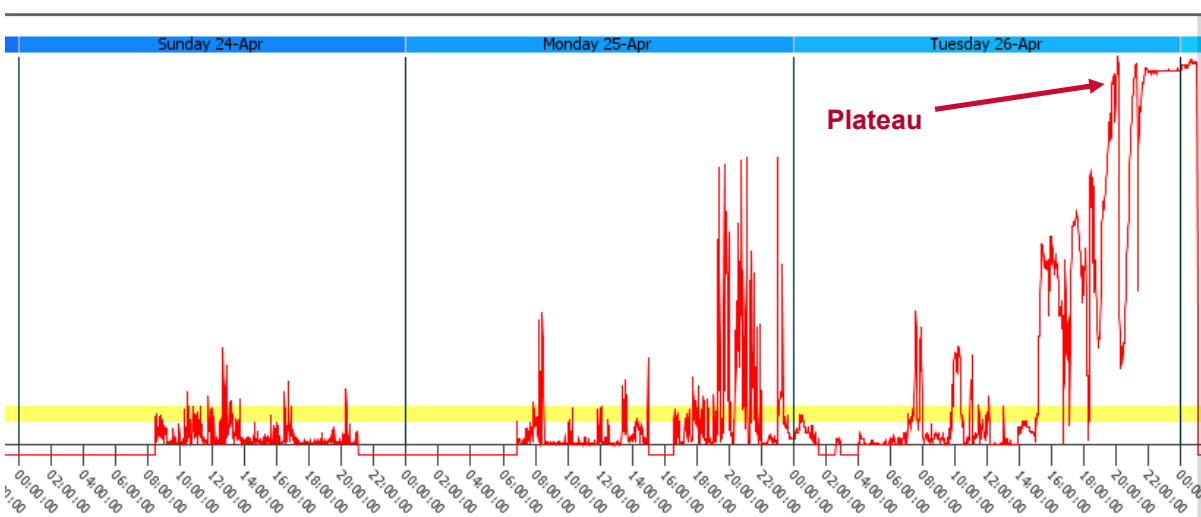
STEP 2 – Establish validity

- Valid (0)
- Not valid (coded 1-3)
  - Spurious (1)
    - Plateauing

Difficult to detect when data is processed



STEP 2 – Establish validity



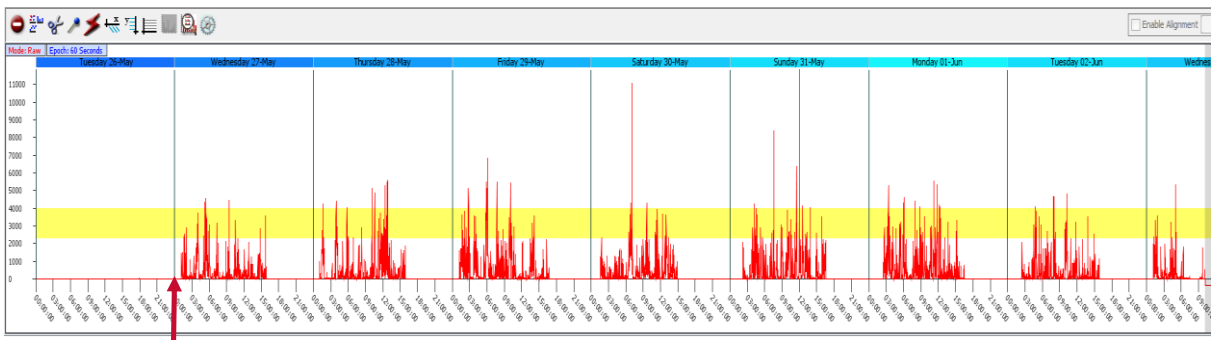
STEP 2 – Establish validity

- Valid (0)
- Non valid (1-3)
  - Spurious (1)
    - Plateauing
    - **Not returning to zero**

STEP 2 – Establish validity

- Valid (0)
- Non valid (1-3)
  - Spurious (1)
    - Plateauing
    - Not returning to zero
  - **Translocated files (2)**

STEP 2 – Establish validity



Mid-night

STEP 2 – Establish validity

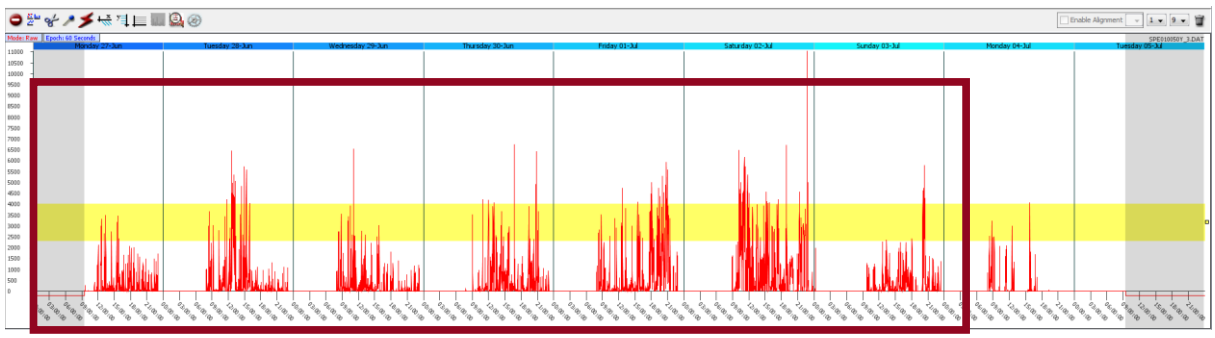
- Valid (0)
- Non valid (1-3)
  - Spurious (1)
    - Plateauing
    - Not returning to zero
  - Translocated files (2)

Valid: ~98%
Spurious: ~1-2%
Translocated: ~0.4%

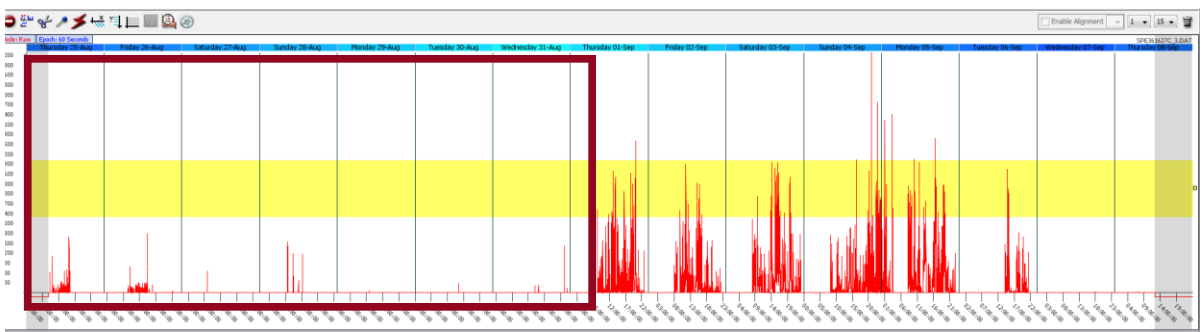
Department of Sports Medicine
NORWEGIAN SCHOOL OF SPORT SCIENCES

Step 3 – Determine wear period

- 7 consecutive days of data are extracted, and read-in file assures best 7 day fit
- Important since every wave of every study was analysed in one batch.



Step 3 – Determine wear period



Applying rules and cut-points

- Non-wear criteria: 60 mins cons strings of zero (2 drops) ALL
- 480 minutes/day default – users can apply own wear time criteria
- Daylight saving time according to study site
- Cut-points and bout-length definitions applied
- **AVAILABLE VARIABLES (>15,000)**

Counts and minutes

- Total, daytype, day, hour-by-hour

Accumulated intensity-specific minutes

- Total, daytype, day, hour-by-hour

Bouted intensity-specific minutes

- Total, day type, day

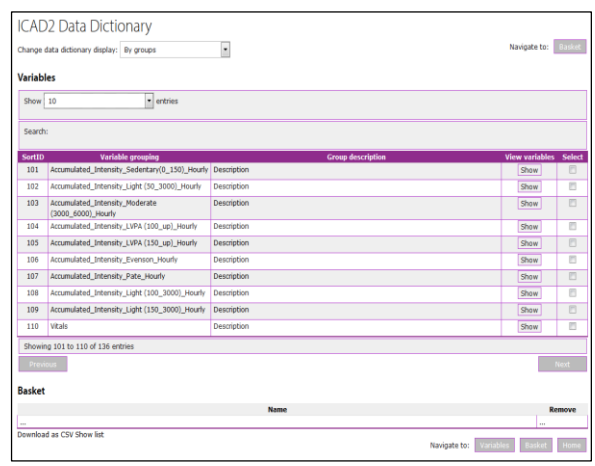
Cutpoint Title	Start Value (>=)	Start Value Type	End Value (<)	End Value Type
0_50	0.0	Counts	50.0	Counts
50_100	50.0	Counts	100.0	Counts
100_150	100.0	Counts	150.0	Counts
150_250	150.0	Counts	250.0	Counts
250_500	250.0	Counts	500.0	Counts
500_750	500.0	Counts	750.0	Counts
750_1000	750.0	Counts	1000.0	Counts
1000_1500	1000.0	Counts	1500.0	Counts
1500_2000	1500.0	Counts	2000.0	Counts
2000_2500	2000.0	Counts	2500.0	Counts
2500_3000	2500.0	Counts	3000.0	Counts
3000_3500	3000.0	Counts	3500.0	Counts
3500_4000	3500.0	Counts	4000.0	Counts
4000_4500	4000.0	Counts	4500.0	Counts
4500_5000	4500.0	Counts	5000.0	Counts
5000_6000	5000.0	Counts	6000.0	Counts
6000_8000	6000.0	Counts	8000.0	Counts
8000_up	8000.0	Counts	99999.0	Counts
Sedentary_50	0.0	Counts	50.0	Counts
Sedentary_100	0.0	Counts	100.0	Counts
Sedentary_150	0.0	Counts	150.0	Counts
LPA_50_2000	50.0	Counts	2000.0	Counts
LPA_100_2000	100.0	Counts	2000.0	Counts
LPA_150_2000	150.0	Counts	2000.0	Counts
LPA_50_3000	50.0	Counts	3000.0	Counts
LPA_100_3000	100.0	Counts	3000.0	Counts
LPA_150_3000	150.0	Counts	3000.0	Counts
MPA_2000_6000	2000.0	Counts	6000.0	Counts
MPA_3000_6000	3000.0	Counts	6000.0	Counts
VPA_6000_up	6000.0	Counts	99999.0	Counts
MVPA_2000_up	2000.0	Counts	99999.0	Counts
MVPA_3000_up	3000.0	Counts	99999.0	Counts
LVPA_50	50.0	Counts	99999.0	Counts
LVPA_100	100.0	Counts	99999.0	Counts
LVPA_150	150.0	Counts	99999.0	Counts
Evenson_SED	0.0	Counts	101.0	Counts
Evenson_LPA	101.0	Counts	2296.0	Counts
Evenson_MPA	2296.0	Counts	4012.0	Counts
Evenson_VPA	4012.0	Counts	99999.0	Counts
Evenson_MVPA	2296.0	Counts	99999.0	Counts
Evenson_LVPA	100.0	Counts	99999.0	Counts
Pate_SED	0.0	Counts	800.0	Counts
Pate_LPA	800.0	Counts	1680.0	Counts
Pate_MPA	1680.0	Counts	3368.0	Counts
Pate_VPA	3368.0	Counts	99999.0	Counts
Pate_MVPA	1680.0	Counts	99999.0	Counts
Pate_lvpa	800.0	Counts	99999.0	Counts

http://www.mrc-epid.cam.ac.uk/research/studies/icad/

NIH Department of Sports Medicine NORWEGIAN SCHOOL OF SPORT SCIENCES

Online data dictionary

- **Summary of data processing**
  - **55,386 accelerometer data files processed**
    - **Around 220 hours of processing time over about 3 weeks**
    - **392 excel spread sheets**
    - **Merged with other variables**
  - **Available through online data dictionary with shopping cart function**



http://www.mrc-epid.cam.ac.uk/research/studies/icad/

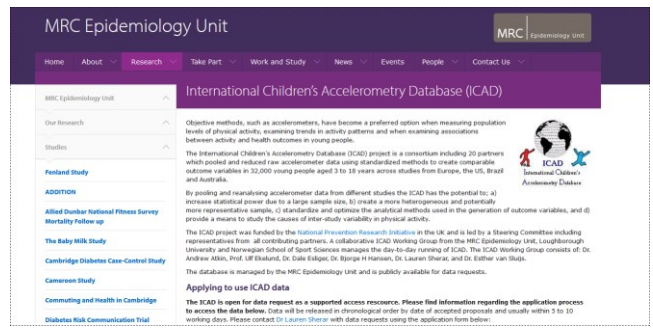
NIH Department of Sports Medicine NORWEGIAN SCHOOL OF SPORT SCIENCES

SortID	Variable grouping	Group description	View variables	Select
101	Accumulated_Intensity_Sedentary(0_150)_Hourly	Description	Show	
102	Accumulated_Intensity_Light (50_3000)_Hourly	Description	Show	
103	Accumulated_Intensity_Moderate (3000_6000)_Hourly	Description	Show	
104	Accumulated_Intensity_LVPA (100_up)_Hourly	Description	Show	
105	Accumulated_Intensity_LVPA (150_up)_Hourly	Description	Show	
106	Accumulated_Intensity_Evenson_Hourly	Description	Show	
107	Accumulated_Intensity_Pate_Hourly	Description	Show	
108	Accumulated_Intensity_Light (100_3000)_Hourly	Description	Show	
109	Accumulated_Intensity_Light (150_3000)_Hourly	Description	Show	
110	Vitals	Description	Show	
111	Accumulated_Intensity_Moderate (2000_6000)_Hourly	Description	Show	
112	Accumulated_Intensity_Vigorous (6000_up)_Hourly	Description	Show	
113	Accumulated_Intensity_MVPA (2000_up)_Hourly	Description	Show	
114	Accumulated_Intensity_MVPA (3000_up)_Hourly	Description	Show	
115	Accumulated_Intensity_LVPA (50_up)_Hourly	Description	Show	
116	Wear_Counts_Daily	Description	Show	
117	Wear_counts_Totals	Description	Show	
118	Weartime_Minutes_Totals	Description	Show	
119	Wear_Counts_Hourly	Description	Show	
120	Wear_Minutes_Hourly	Description	Show	
121	StD_Age	Description	Show	
122	AM_Birthweight	Description	Show	
123	AM_Glucose	Description	Show	
124	AM_Insulin	Description	Show	
125	AM_HDL	Description	Show	

119	Wear_Counts_Hourly	Description WearCtsFriHr03 WearCtsFriHr04 WearCtsFriHr05 WearCtsFriHr06 WearCtsFriHr07 WearCtsFriHr08 WearCtsFriHr09 WearCtsFriHr10 WearCtsFriHr11 WearCtsFriHr12 WearCtsFriHr13 WearCtsFriHr14 WearCtsFriHr15 WearCtsFriHr16 WearCtsFriHr17 WearCtsFriHr18	Hide	
120	Wear_Minutes_Hourly	Description WearMinFriHr03 WearMinFriHr04 WearMinFriHr05 WearMinFriHr06 WearMinFriHr07 WearMinFriHr08 WearMinFriHr09 WearMinFriHr10 WearMinFriHr11 WearMinFriHr12 WearMinFriHr13 WearMinFriHr14 WearMinFriHr15 WearMinFriHr16 WearMinFriHr17 WearMinFriHr18	Hide	

Online data dictionary

- **ICAD is an open data base**
- **Data base managed by the MRC Epidemiology Unit**
- **Simple Application procedure**
- **Data user agreement**
- **Authorship rules/Publication guidelines**
- **Working group**
- **Steering Group**
- **Newsletter**



Department of Sports Medicine
NORWEGIAN SCHOOL OF SPORT SCIENCES

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)

Andrew J. Atkin[1,2*], Stuart J. H. Biddle[3], Stephanie T. Broyles[4], Mai Chinapaw[5], Ulf Ekelund[6], Dale W. Esliger[7], Bjorge H. Hansen[6], Susi Kriemler[8], Jardena J. Puder[9], Lauren B. Sherar[7], Esther M. F. van Sluijs[2] and On behalf of the International Children's Accelerometry Database (ICAD) Collaborators

Conclusion: The project to expand ICAD further demonstrates the feasibility of pooling data on physical activity, sedentary behaviour and potential determinants from multiple studies. Key to this process is the rigorous conduct and reporting of retrospective data harmonisation, which is essential to the appropriate analysis and interpretation of derived data. These documents, made available through the ICAD website, may also serve as a guide to others undertaking similar projects.

(Atkin et al, IJBNPA 2017)

Department of Sports Medicine
NORWEGIAN SCHOOL OF SPORT SCIENCES

Cross-Sectional Associations of Reallocating Time Between Sedentary and Active Behaviours on Cardiometabolic Risk Factors in Young People: An International Children's Accelerometry Database (ICAD) Analysis

Bjørge Herman Hansen[1] · Sigmund Alfred Anderssen[1] · Lars Bo Andersen[1,2] · Maria Hildebrand[1] · Elin Kolle[1] · Jostein Steene-Johannessen[1] · Susi Kriemler[3] · Angie S. Page[4] · Jardena J. Puder[5] · John J. Reilly[6] · Luis B. Sardinha[7] · Esther M. F. van Sluijs[8] · Niels Wedderkopp[9] · Ulf Ekelund[1] · On behalf of the International Children's Accelerometry Database (ICAD) Collaborators

Key Points

Our results show beneficial theoretical associations between replacing as little as 10 min/day of sedentary time with an equal amount of time spent in moderate-to-vigorous physical activity and a wide array of cardiometabolic risk markers in healthy youth.

Replacing sedentary time with an equal amount of light physical activity showed minor beneficial associations with cardiometabolic risk markers.

Replacing sedentary time with active behaviours, particularly those of at least moderate intensity, appears to be an effective strategy to reduce cardiometabolic risk in young people.

(Hansen et al, Sports Med 2018)

ORIGINAL ARTICLE

Does adiposity mediate the relationship between physical activity and biological risk factors in youth?: a cross-sectional study from the International Children's Accelerometry Database (ICAD)

J Tarp[1], A Bugge[1], LB Andersen[2,3], LB Sardinha[4], U Ekelund[3,5], S Brage[5] and NC Møller[1] On behalf of the International Children's Accelerometry Database (ICAD) Collaborators

CONCLUSIONS: One hour of daily moderate-to-vigorous physical activity was associated with clinically relevant differences in metabolic control compared to engagement in less than this minimally recommended amount. The majority of the difference was explained by the direct effect of physical activity.

(Tarp et al, Int J Obes 2017)

NIH Department of Sports Medicine NORWEGIAN SCHOOL OF SPORT SCIENCES

Association between maternal education and objectively measured physical activity and sedentary time in adolescents

Lauren B Sherar,[1] Tom P Griffin,[2] Ulf Ekelund,[3,4] Ashley R Cooper,[5] Dale W Esliger,[1] Esther M F van Sluijs,[4,6] Lars Bo Andersen,[3,7] Greet Cardon,[8] Rachel Davey,[9] Karsten Froberg,[7] Pedro C Hallal,[10] Kathleen F Janz,[11] Katarzyna Kordas,[12] Susi Kriemler,[13] Russell R Pate,[14] Jardena J Puder,[15] Luis B Sardinha,[16] Anna F Timperio,[17] Angie S Page[5]

Conclusions Across a number of international samples, adolescents of mothers with lower education may not be at a disadvantage in terms of overall objectively measured PA.

(Sherar et al, J Epi Com Health 2016)

NIH Department of Sports Medicine NORWEGIAN SCHOOL OF SPORT SCIENCES

MEET THE TEAM

ICAD Working Group
Ulf Ekelund
Esther van Sluijs
Lauren Sherar
Dale Esliger
Andy Atkin
Bjørge Herman Hansen

ICAD Data management
Susie Boatman
Adam Dickinson
Inge Loudon
Tom White
Anna Melachrou

ICAD Partners
Katarzyna Kordas/Lynn Molloy/Kate Northstone
Jardena J Puder
Greet Cardon
Rachel Davey
Russ Pate
Jo Salmon
Lars Bo Andersen
Niels Christian Moller
Luis B Sardinha
Sigmund Anderssen
Anna Timperio
Kathleen F Janz
Susi Kriemler
John J Reilly
Angie Page
Pedro C Hallal

MRC Epidemiology Unit | Loughborough University

NIH Department of Sports Medicine NORWEGIAN SCHOOL OF SPORT SCIENCES