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

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

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
  • Population levels and cultural differences
  • Dose-response:
    • Frequency, Intensity, Duration, Mode
  • Inter-relations between movement behaviours, sedentary behaviour and sleep)
• What we didn`t know
  • Effect modifiers: population subgroups
  • Correlates and Determinants
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
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
International children’s accelerometry database (ICAD): Design and methods

Lauren B Sherar¹, Pippa Grieve², Dale W Esliger¹, Ashley R Cooper³, Ulf Ekelund⁴⁵, Ken Judge⁶ and Chris Riddoch⁶

- 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)
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
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)

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
- Reintegrated 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)
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

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

• Translocated files (2)
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%
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.
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

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

---

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/
<table>
<thead>
<tr>
<th>Source</th>
<th>Variable / Group</th>
<th>Group description</th>
<th>View variables</th>
<th>Select</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>Accumulated_Intensity_Sedentary [8, 150] Hourly</td>
<td>Description</td>
<td>Show</td>
<td></td>
</tr>
<tr>
<td>102</td>
<td>Accumulated_Intensity_Light [50, 3000] Hourly</td>
<td>Description</td>
<td>Show</td>
<td></td>
</tr>
<tr>
<td>103</td>
<td>Accumulated_Intensity_Moderate [3000, 5000] Hourly</td>
<td>Description</td>
<td>Show</td>
<td></td>
</tr>
<tr>
<td>104</td>
<td>Accumulated_Intensity_VPA [100, up] Hourly</td>
<td>Description</td>
<td>Show</td>
<td></td>
</tr>
<tr>
<td>105</td>
<td>Accumulated_Intensity_LVPA [150, up] Hourly</td>
<td>Description</td>
<td>Show</td>
<td></td>
</tr>
<tr>
<td>106</td>
<td>Accumulated_Intensity_Extremes Hourly</td>
<td>Description</td>
<td>Show</td>
<td></td>
</tr>
<tr>
<td>107</td>
<td>Accumulated_Intensity_Pale Hourly</td>
<td>Description</td>
<td>Show</td>
<td></td>
</tr>
<tr>
<td>108</td>
<td>Accumulated_Intensity_Light [100,5000] Hourly</td>
<td>Description</td>
<td>Show</td>
<td></td>
</tr>
<tr>
<td>109</td>
<td>Accumulated_Intensity_Light [150,2000] Hourly</td>
<td>Description</td>
<td>Show</td>
<td></td>
</tr>
<tr>
<td>110</td>
<td>Vitalb</td>
<td>Description</td>
<td>Show</td>
<td></td>
</tr>
<tr>
<td>111</td>
<td>Accumulated_Intensity_Moderate [2000,5000] Hourly</td>
<td>Description</td>
<td>Show</td>
<td></td>
</tr>
<tr>
<td>112</td>
<td>Accumulated:Intensity_Vigorous [5000, up] Hourly</td>
<td>Description</td>
<td>Show</td>
<td></td>
</tr>
<tr>
<td>113</td>
<td>Accumulated_Intensity_VPA [2000, up] Hourly</td>
<td>Description</td>
<td>Show</td>
<td></td>
</tr>
<tr>
<td>114</td>
<td>Accumulated_Intensity_VPA [3000, up] Hourly</td>
<td>Description</td>
<td>Show</td>
<td></td>
</tr>
<tr>
<td>115</td>
<td>Accumulated:Intensity_LVPA [50, up] Hourly</td>
<td>Description</td>
<td>Show</td>
<td></td>
</tr>
<tr>
<td>116</td>
<td>Wear_Counts_Daily</td>
<td>Description</td>
<td>Show</td>
<td></td>
</tr>
<tr>
<td>117</td>
<td>Wear_counts_Totals</td>
<td>Description</td>
<td>Show</td>
<td></td>
</tr>
<tr>
<td>118</td>
<td>Weartime_Minutes_Totals</td>
<td>Description</td>
<td>Show</td>
<td></td>
</tr>
<tr>
<td>119</td>
<td>Wear_Counts_Hourly</td>
<td>Description</td>
<td>Show</td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>Wear_Minutes_Hourly</td>
<td>Description</td>
<td>Show</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>Variable / Group</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>119</td>
<td>Wear_Counts_Hourly</td>
<td>Description</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>Wear_Minutes_Hourly</td>
<td>Description</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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

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, Stuart J. H. Biddle, Stephanie T. Broyles, Mai Chinapaw, Ulf Ekeliund, Dale W. Eslinger, Bjorge H. Hansen, Susi Kriemler, Jardena J. Puder, Lauren B. Sherar, Esther M. F. van Sluijs 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.
Cross-Sectional Associations of Realocating Time Between Sedentary and Active Behaviours on Cardiometabolic Risk Factors in Young People: An International Children’s Accelerometry Database (ICAD) Analysis

Bjørge Herman Hansen¹, Sigmund Alfred Andersen¹, Lars Bo Andersen¹,², Maria Hildebrand¹, Else Kolle¹, Jostein Steine-Johannessen¹, Susi Kylenler³, Angie S. Page⁴, Jordana J. Puder⁵, John J. Reilly⁶, Luis R. Sardinha⁷, Esther M. F. van Sluijs⁸, Niels Pedderkopp⁹, Ulf Ekelund¹, 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 Tarp¹, A Bugge¹, LB Andersen²,³, LB Sardinha⁴, U Ekelund⁵,⁶, S Brage² and NC Moller¹ 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)
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,6,7 Lars Bo Andersen,4,7 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 Page5

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)