COLLECTING ACTIVITY AND SLEEP DATA IN THE HEALTH AND RETIREMENT STUDY

Jessica D. Faul

JPSM MPSM Seminar
April 7, 2021
MOTIVATION FOR COLLECTING ACTIGRAPHY DATA

• Actigraphy – non-invasive collection of physical movement and sleep data to supplement self-reports

• Research has shown that self-reports produce much smaller age differences and cross-national differences in activity levels than those measured with actigraphy (Kapteyn et al. 2017; Kapteyn et al. 2018)

• Self-report alone might not be adequate to evaluate group differences

Kapteyn et al. (2018). Epidemiology and Community Health
ACTIGRAPHY IN POPULATION-BASED STUDIES

• Use of accelerometers more feasible now in large-scale epidemiological and population-based studies (lower cost, longer battery life)

• Several have or are planning to test or use these devices to collect objective data to relate to clinical endpoints

• Women’s Health Study, UKBiobank, English Longitudinal Survey of Ageing (ELSA), The Irish Longitudinal Study on Ageing (TILDA), The Survey of Health, Ageing and Retirement in Europe (SHARE), Midlife in the U.S (MIDUS)...

HRS ACTIGRAPHY PILOT

• HRS was asked to conduct a feasibility pilot of actigraphy, not to implement at scale

• The goal of the pilot was to test our protocol for using activity monitors to collect physical activity and sleep data from HRS respondents as well self-reported information via a short questionnaire

• Pilot designed to inform us about response rates and logistics and about processing and interpreting data
CHALLENGES WE KNEW GOING INTO THIS...

• More Data!
  • As battery and memory increased, it became possible to collect and store high-throughput, three-axial, frequently measured acceleration data for several days-weeks
  • Supplementary sensors including gyroscopes, thermometers, inclinometers, pulsometers, light intensity and skin conductance sensors

• Data Analytics
  • Activity/sleep measures are summaries of the raw data - depend on the device manufacturer, software version, and body location
  • Proprietary aggregated measures from companies - frequent updates to algorithm software (version control issues)
  • Need development of well-defined, open-source, reproducible data summary tools

• Harmonization

**Figure 3** Acceleration from three orthogonal axes of an accelerometer located on the hip (left column) and left wrist (right column), while dealing cards (top row), getting dressed (middle row) and walking (bottom row). Each axis data are shown in a different color.
ACTIGRAPHY PRE-PILOT

- **Decide which device** - To prepare for the pilot we conducted pre-pilot work to test various wrist devices (Crimmins, Kapteyn)

- **Test settings** - Conducted additional testing of selected device to test various setting options (Faul/Mitchell/Guimaraes/Troupe)
ACTIGRAPHY DEVICES

- PrePilot work compared Fitbit and a GENEactiv mid-range model to Actigraph

<table>
<thead>
<tr>
<th>Fitbit Flex</th>
<th>GENEActiv Original</th>
<th>Actigraph wgt3x</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Fitbit Flex" /></td>
<td><img src="image2.png" alt="GENEActiv Original" /></td>
<td><img src="image3.png" alt="Actigraph wgt3x" /></td>
</tr>
</tbody>
</table>
# ACTIGRAPHY DEVICES

<table>
<thead>
<tr>
<th>Fitbit</th>
<th>GENEActiv</th>
<th>Actigraph</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easily wearable (higher compliance)</td>
<td>Somewhat wearable</td>
<td>Difficult to wear, “ugly”</td>
</tr>
<tr>
<td>&lt; $100</td>
<td>$200 (mid-level model)</td>
<td>$400</td>
</tr>
<tr>
<td>Data uploaded to 3rd party</td>
<td>Control data stream</td>
<td>Control data stream</td>
</tr>
<tr>
<td>Syncing required (requires IWER to assist and problems occur with some frequency)</td>
<td>No syncing (can be mailed and activated by user)</td>
<td>No syncing (can be mailed and activated by user)</td>
</tr>
<tr>
<td>Water resistant</td>
<td>OK in shower</td>
<td>OK in shower</td>
</tr>
<tr>
<td>Short battery life (3 days)</td>
<td>Recording at 100Hz lasts for 7 days (at least) and 10Hz records for 45 days</td>
<td>100Hz + Sleep mode = 10 days 60Hz + Sleep mode = 14 days</td>
</tr>
</tbody>
</table>
ACTIGRAPHY DEVICES

• HRS originally planned on using the Fitbit device – preferred by older respondents (Mercer et al. 2016) and thought about allowing respondents to keep the device as an incentive

• Initial pilot testing of the Fitbit device against the GENEActiv device used by A. Kapteyn and ELSA within a volunteer sample (Crimmins) showed that results could not be harmonized easily

• A second prepilot of GENEActiv, Actigraph and the iPhone (only examining activity not sleep) showed that with comparable settings (Hz) data could be well-harmonized (higher Hz translates into higher measurement frequency)

VM - Vector Magnitude is the square root of the sum of the squares of each axis of data \((x,y,z)\) – used to create cut-points to classify physical activity

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>iPhone and GENEActiv at 60 Hz</td>
<td>0.986</td>
</tr>
<tr>
<td>iPhone and GENEActiv at 100 Hz</td>
<td>0.916</td>
</tr>
<tr>
<td>iPhone and Actigraph at 60 Hz</td>
<td>0.864</td>
</tr>
<tr>
<td>GA-60Hz and GA-100Hz</td>
<td>0.892</td>
</tr>
<tr>
<td>GA-60Hz and Actigraph-60Hz</td>
<td>0.879</td>
</tr>
<tr>
<td>GA-100Hz and Actigraph-60Hz</td>
<td>0.693</td>
</tr>
</tbody>
</table>

Kapteyn et al. – Pilot work
# ACTIGRAPHY DEVICES

<table>
<thead>
<tr>
<th></th>
<th>Fitbit</th>
<th>GENEActiv</th>
<th>Actigraph</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easily wearable (higher compliance)</td>
<td>Easily wearable (higher compliance)</td>
<td>Somewhat wearable</td>
<td>Difficult to wear, “ugly”</td>
</tr>
<tr>
<td>&lt; $100</td>
<td>&lt;$100</td>
<td>$200 (mid-level model)</td>
<td>$400</td>
</tr>
<tr>
<td>Data uploaded to 3rd party</td>
<td>Data uploaded to 3rd party</td>
<td>Control data stream</td>
<td>Control data stream</td>
</tr>
<tr>
<td>Syncing required (requires IWER to assist and problems occur with some frequency)</td>
<td>Syncing required (requires IWER to assist and problems occur with some frequency)</td>
<td>No syncing (can be mailed and activated by user)</td>
<td>No syncing (can be mailed and activated by user)</td>
</tr>
<tr>
<td>Water resistant</td>
<td>Water resistant</td>
<td>OK in shower</td>
<td>OK in shower</td>
</tr>
<tr>
<td>Short battery life (3 days)</td>
<td>Short battery life (3 days)</td>
<td>Recording at 100Hz lasts for 7 days (at least) and 10Hz records for 45 days</td>
<td>100Hz + Sleep mode =10 days 60Hz + Sleep mode =14 days</td>
</tr>
</tbody>
</table>
GENEACTIV ORIGINAL

- GENEActiv used by ELSA, TILDA, Whitehall, NSHAP (2020)
- Durable and waterproof
- Records light and uses a temperature sensor to determine whether it has been taken off the wrist
  - Light measurement is critical in helping to determine when the respondent is in bed and trying to go to sleep (i.e., the rest interval), which is necessary for scoring sleep parameters
- Produces usable report with minutes of activity and sleep
- GENEActiv raw data (m/s² or gravitational acceleration for all 3 dimensions) can be processed with an R package
- Provides estimate of steps
PRE-PILOT – TESTING GENEACTIV SETTINGS

• Spring 2019
• Volunteers (college students) each wore 2 devices for 7 days
• 3 Groups:
  • Control (n=8) – 2 identical devices
  • Treatment 1 (n=7) – with/without preloaded data entered during device set-up (height, weight, handedness)
  • Treatment 2 (n=8) – with/without light sensor covered
• Evaluated mins sleep, sedentary, light/moderate/vigorous activity averaged over 7 days
• Paired T Tests (whether difference between watches=0)
PRE-PILOT – TESTING GENEACTIV SETTINGS

• Sleep – no significant differences (at p < 0.05) but both Tx watches (with preloaded data and with covered light sensor) in general recorded more sleep (~.12 SD more) – approx 7 more mins per night

• Sedentary Activity – no significant differences (at p < 0.05) but Tx 1 Preload watch ~.5 SD fewer mins (p=0.35)
## PRE-PILOT – TESTING GENEACTIV SETTINGS

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean Diff (mins)</th>
<th>SE Diff</th>
<th>Paired T Test P != 0</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Light Activity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>8</td>
<td>13.13</td>
<td>13.41</td>
<td>0.360</td>
</tr>
<tr>
<td>Preload</td>
<td>7</td>
<td>-12.00</td>
<td>14.43</td>
<td>0.438</td>
</tr>
<tr>
<td>Light</td>
<td>8</td>
<td>19.25</td>
<td>9.28</td>
<td>0.077</td>
</tr>
</tbody>
</table>

**Control Group – No differences**

Tx1 (Preload) – Watch preloaded with height, weight, handedness consistently recorded more activity than control = avg 13 more minutes moderate activity per day

Tx2 (Light) – Watch with light sensor covered recorded fewer mins activity
PRE-PILOT – TESTING GENEACTIV SETTINGS

• Handedness significantly impacts summary report data
• HRS decided to not preload handedness in the 2019 pilot but collected and entered on which wrist the device was worn and handedness at data download and summarization
HRS ACTIGRAPHY PILOT: KEY PROTOCOL DECISIONS

• Geneactiv wrist device (same as ELSA, UAS)
• 10 days of wear so both weekdays and weekends are covered (trying to get 7 useable days)
• 50 Hz
• Eligible sample + controls selected prior to HRS 2018 interviews
• Developed a short questionnaire on activity/sleep during study duration (using questions from the Consensus Sleep Diary (Carney, 2012) and Pittsburgh Sleep Quality Index (PSQI))
• Sample divided into 4 batches so devices could be re-used and to balance work-flow
HRS ACTIGRAPHY PILOT: QUESTIONS ADDED TO THE 2018 CORE SURVEY

• We anticipated concerns about self-selection into completing the pilot and possible behavioral change due to participation in the pilot.

• As part of the overall pilot strategy, the selected respondents (eligibles & controls) were asked additional questions in 2018 core interview about sleep and activity.
HRS ACTIGRAPHY PILOT: QUESTIONS ADDED TO THE 2018 CORE SURVEY

Sleep:

- What time do you usually go to bed and start trying to fall asleep on weekdays or work days?
- What time do you usually go to bed and start trying to fall asleep on weekends?
- What time do you usually wake up on weekdays or work days?
- What time do you usually wake up on weekends?

[ENTER HOUR AND MINUTE AND AM/PM]
HRS ACTIGRAPHY PILOT: QUESTIONS ADDED TO THE 2018 CORE SURVEY

Activity:

• About how much time do you usually spend on your feet – for example, standing, walking, etc. – on a typical day? This can include time at work, at home or somewhere else, or getting to and from places.

• How much time do you usually spend sitting on a typical day? Again, this can include time at work, at home or somewhere else, getting to and from places, or with friends or family. Do not include time spent sleeping.

[ENTER MINUTES OR HOURS]
Mail: Eligible respondents sent invitation letter, informed consent, a $25 token of appreciation, pre-paid return envelope

Reminder postcard was sent 2 weeks later

Upon receipt of consent, an enrollment letter was sent to respondents along with an activity monitor, a brief questionnaire and sleep diary, instructions for using the device, and a pre-addressed, pre-paid return envelope

R completes device wearing/q’naire completion, returns them

If the device and questionnaire were not received within approx. 2 weeks of the expected return date (based on the 10-day measurement period plus 3 days expected for return following this period), a device and questionnaire return reminder postcard were sent

R receives additional $25 and a thank you letter with results letter (showing average sleep and activity results from the 10-day wear trial) or letter indicating that results were not reportable when device returned

Follow up call if 1 month passes and device is not returned
### YOU LOSE SOME AT EACH STEP

<table>
<thead>
<tr>
<th>Step</th>
<th>N</th>
<th>progress rate at each step</th>
</tr>
</thead>
<tbody>
<tr>
<td>eligible</td>
<td>380</td>
<td></td>
</tr>
<tr>
<td>consent</td>
<td>204</td>
<td>53.7%</td>
</tr>
<tr>
<td>refused</td>
<td>64</td>
<td>16.8%</td>
</tr>
<tr>
<td>no answer</td>
<td>112</td>
<td>29.5%</td>
</tr>
<tr>
<td>sent device</td>
<td>220</td>
<td>100%</td>
</tr>
<tr>
<td>returned</td>
<td>183</td>
<td>83.2%</td>
</tr>
<tr>
<td>complete data</td>
<td>151</td>
<td>82.5%</td>
</tr>
<tr>
<td>net RR</td>
<td>48%</td>
<td>(complete/eligible)</td>
</tr>
</tbody>
</table>

420 Selected, 1 deceased and 39 removed because Spanish-speaking

About 15% of people fail to return device within a month. This could be a significant cost at large scale to recover or write off

About 17% of cases had no data (device set-up failure) or too little data (wear failure)
HRS ACTIGRAPHY PILOT: PRELIMINARY FINDINGS

- Selection into the pilot
- Data validation
- Did participation change behavior?
- Did getting a report change behavior among those who participated?
OLDER RESPONDENTS LESS LIKELY TO PARTICIPATE

Consent  Refused  NR

50-59
69-69
70-79
80+

Consent  Refused  NR
NO LARGE DIFFERENCES BY SEX

Male
- Consent: 60%
- Refused: 10%
- NR: 30%

Female
- Consent: 65%
- Refused: 15%
- NR: 20%
LOWER CONSENT AMONG NH BLACKS, HIGHER NR
NOT HUGE DIFFERENTIALS BY SELF-RATED HEALTH

- Excellent
- Very Good
- Good
- Fair
- Poor

Legend:
- Consent
- Refused
CONSENTERS REPORT MORE VIGOROUS ACTIVITY

Mean NAGI (0-10)
Consent = 1.4
Refused = 2.3
NR = 2.2
CONSENTERS REPORT LESS TROUBLE SLEEPING

- Most of the time
- Sometimes
- Rarely or Never

Consent
Refused
NR
DATA APPEAR TO HAVE SOME VALIDITY: DAILY MINUTES OF ACTIVITY (DEVICE) BY INTENSITY AND AGE

- Light
- Moderate
- Vigorous

Age Groups:
- 50-59
- 60-69
- 70-79
- 80+
DATA APPEAR TO HAVE SOME VALIDITY:
DAILY MINUTES OF SLEEP (DEVICE) BY AGE
DATA APPEAR TO HAVE SOME VALIDITY: MEAN DAILY MINUTES OF LIGHT AND MODERATE ACTIVITY BY SELF-REPORTED ACTIVITY
DID PARTICPATION CHANGE BEHAVIOR – CHANGES FROM 2018 TO 2020?

- Completed Cases vs Controls
- Changes in answers between 2018 and 2020 were quantified as 1 (change in positive direction), 0 (no change) or -1 (change in undesirable direction) for the following variables
- No significant difference between Completed Cases and Controls on change in self-rated health, trouble falling asleep, waking in the night, waking too early, feeling rested, activity levels
DID GETTING A REPORT CHANGE BEHAVIOR AMONG THOSE WHO PARTICIPATED?

• Completed cases who received a results letter vs those that did not

• As above, changes in answers between 2018 and 2020 were quantified as 1 (change in positive direction), 0 (no change) or -1 (change in undesirable direction) for the following variables

• No significant difference between Completed Cases and Controls on change in self-rated health, trouble falling asleep, waking in the night, waking too early, feeling rested, activity levels

* although note small sample size of those that did not receive a letter (n=32)
The GENEActiv devices can be difficult to charge and access via the software provided and the lack of visibility can make it difficult to ensure the devices are set up correctly
- Makes rapid, bulk efforts difficult
- Some devices returned with no data available to process due to set-up error

Downloading and processing data was a major challenge – download and summary is time consuming (several hours per device, max 2 at a time) – would need more computing capacity

GENEActiv changed data summarization algorithms between the Prepilot (April) and Pilot (June) – now using an R markdown file
- Advantage: R returns total steps as a summary measure
- Disadvantages: summary statistics not the same across algorithms, need to carefully consider this if you want to harmonize with other groups

Operationally, the mailing process has been efficient, easy to track
HRS ACTIGRAPHY PILOT: FIELD EXPERIENCE

• Due to the mode used for this study (mail), some loss was expected. After consenting to the study and a device being sent, 16 respondents did not return their activity monitors.

• Don’t underestimate turnaround time to prep a returned device for mailing – staggered mailings.

• Due to the design of the watchbands, cleanliness became a priority in between mailing batches. Each band contained a series of grooves, which would collect skin particles during use.

• Additionally, respondents found the lack of display confusing during use (couldn’t tell whether it was “on”)


Anecdotally, respondents were receptive to the study and we received positive feedback by email, letters, and phone calls.

Those who have agreed to participate seem excited to try a new method of data collection.
HRS ACTIGRAPHY PILOT: DATA CHALLENGES

• Need to consider how/what data products would HRS produce for release
• Summary measures (time spent sedentary/light/moderate/vigorous activity, minutes sleeping, steps) per day – could be public data
• Summarizing sleep data is tricky and best done in conjunction with sleep diary info
• Raw data files are large and require expertise to analyze
CONCLUSIONS

• Increasingly possible and affordable to collect activity and sleep data in large-scale studies

• Many of the devices that are optimized to collect activity data are not ideal for sleep (supplement with self-reported sleep info)

• Significant resources needed to download and manage data collected on large scale

• Pre-processing pipelines still need to take into account device selection and calibration, sampling frequency, missing data

• Selection issues so likely can’t replace self-reported data entirely

• Consent would likely improve if asked FTF not by mail

• Relationship to long term outcomes still TBD
HARMONIZATION ISSUES

• Possible to compare and combine studies that collect raw accelerometry data at the same location on the body using open source summary algorithms that are increasingly available (further development needed)

• Still need to harmonize device and pre-processing decisions as much as possible

• Harmonize survey content as well

• Release raw data as well as summarized files harmonized across cohorts using identical algorithm software

• Consider data users
  • Parallel with genetic data in terms of considerations for use - most will not be experienced with processing raw files, high performance computing environments

• Potential confounding of sample characteristics and measurement context (e.g. season) not stable across field period
THANK YOU TO A GREAT TEAM!

Survey Research Operations unit (SRO):
• Daniel Tomlin
• Rick Krause
• Ashwin Dey
• Tim Wright
• Anna Fuqua Smith

HRS Data Entry:
• Gavin Kumpelis
• Amina Troupe

UROP Students:
• Giovana Guimaraes
• Amina Troupe
THANK YOU!

• Visit our website: hrsonline.umich.edu
ADDITIONAL RESEARCH NEEDED

• Comparison of results from different actigraphy devices and the variety of algorithms used to evaluate actigraphy data in order to further establish standards of actigraphy technology

• Research is needed to establish standards for setting start and stop times of the sleep and wake periods (might be device specific)

• More research is needed to assess the reliability of actigraphy under various circumstances, and to determine what parameters may be used to assess the quality of actigraphic data

• There is a need for well-designed studies that include technical details related to the administration and scoring of actigraphy - whether visual inspection of data is performed, how missing data is handled etc.