Description of files for raw and derived data for the interventional clinical study

Contact/correspondence for data set: Arne Mueller (arne.mueller@novartis.com)

Access/URLs:

- • Independent validation study
  - o Derived data and metadata: https://doi.org/10.5281/zenodo.2841297
- • Interventional clinical trial
  - o Derived and metadata: https://doi.org/10.5281/zenodo.2846013
- • Both studies
  - o Full dataset: https://mueller-et-al-2019.s3.amazonaws.com/index.html

General notes:

Note, the device records into one or more data files, but these files belong to the same recording. Also note, that patient usually has two belts (one leather belt and one comfortable flex belt) with one device each and decides which belt he/she wants to wear when. If a belt is open (belt buckle not closed) the belt records a belt open flag. However, a belt might be stored unworn but in a closed state.

Note, in actibelt (when worn correctly) x, y and z values correspond to vertical, lateral and back-forward acceleration.

Derived

meta.csv: meta data to map patients to recoding files

1. belt.id: serial number of the device (repeated if the device was re-used)
2. start: timestamp in UTC when the recording was started
3. end: timestamp in UTC when the recording was ended
4. duration: duration of the recording in seconds
5. stopcode: stop code at the end of a recording (vendor specific)
6. format: version of the proprietary data format
7. file: name of a raw file (several files can be present for one UUID)
8. uuid: unique ID per belt, patient and recording period. Note one UUID can have several files
9. SUBJID: anonymized patient identifier
10. gender: gender of patient
11. patient_timezone: time zone of the patient to derive local time for start and end of recording
12. range_of_birth: a range of 5 years for the patient's year of birth (is left inclusive and right exclusive, e.g. (1930-1935] is >= 1930 and < 1935)

clin_acti_walk_tests.csv: Clinical walk test during which the actibelt was worn

1. SUBJID: anonymized patient id
2. VISITNUM: visit number (coding according to internal use)

3. repeat_in_visit: test repetition within visit
4. test_date: date when the test was performed
5. watch_speed: walking speed calculated by stop watch and distance
6. acti_speed: estimated speed by actibelt
7. uuid: UUId of actibelt measurement
8. file: actibelt filename
9. patient_timezone: time zone of measurement
10. test_start_time: start timestamp of test in UTC
11. test_stop_time: end timestamp of test in UTC
12. test: The walk test conducted (4mWT: 4 meter, 6MWT: six minute, 400mWT: 400 meter)

daily-summaries.csv: daily step, wear time and activity summaries

1. SUBJID: aanonymized patient id
2. day: day/date for this summary
3. VISITNUM: internal visit number coding
4. daytime: daytime of the day (night >20h to 4h, from or daytime >4h to 20h)
5. isvisit: logical, TRUE is this day was a clinical visit day
6. deltaday: the number of days to the nearest visit
7. total_weartime: hours of wearing during this time
8. max_activity: the maximum hourly activity count for this time
9. total_activity: sum of activity count for this time
10. total_steps: number of steps in this time

weartime.csv: wear time per patient within each day and hour of day

bodyheight-femur.csv: patient height and length of femur in cm

steps.h5: Steps for each patient in HDF5 format

1. SUBJID: anonymized patient id
2. time: time of heel strike (UTC)
3. running: was this a running step (logical)?
4. speed: estimated step speed in m/sec
5. duration: estimated duration of the step in sec.
6. distance: estimated distance of the step in meters
7. side: left or right foot (estimated)?
8. bout: the bout identifier this step is in
9. freq: windowed step frequency
10. walk.ratio: windowed walking ratio
11. belt: the belt index this step was estimated from (for internal use)
12. tz: time zone of this step (patient time zone)
13. daytime: categorized daytime of the time stamp

bout-summaries.h5: steps summarized into bouts per hour and

1. SUBJID: anonymized patient id
2. VISITNUM: internal visit number coding
3. bout: bout identifier
4. hour: hour in day
5. days:
6. day: day/date
7. daytime: categorized daytime
8. deltaday: number of days to the nearest visit
9. Steps: number of steps in bout
10. avg_duration: mean step duration in bout in sec.
11. avg_distance: mean step distance in bout in meters
12. total_distance: total distance covered with this bout (meters)
13. total_duration: total duration of this bout in sec.
14. med_speed: median step speed in this bout in m/sec
15. avg_speed: mean step speed in this bout in m/sec
16. min_speed: minimal step speed in this bout in m/sec
17. max_speed: maximum step speed in this bout in m/sec
18. q25_speed: 25th percentil step speed in this bout in m/sec
19. q75_speed: 75th step speed in this bout in m/sec
20. sd_speed: step speed standard deviation this bout
21. iqr_speed: step speed inter-quartile range of this bout
22. avg_walkratio0: mean walk ratio of steps in this bout
23. avg_freq: average windowed step frequency in this bout as reported by step detection software
24. Nsteplength: body hight-normalised step length (mean(distance*1000)/height*mean_height)
25. Ncadence: mean(60/duration)*sqrt(height/mean_height)
26. avg_walkratio: nsteplength/ncadence
27. isvisit: is this day a visit (logical)?
28. total_weartime: total daytime weartime
29. bucket: categorized bout length in number of steps

Raw

Each UUID (see meta.csv) is an HDF5 file with each file recording for a dataset for which the x, y and z acceleration is stored as well as the KSS with 100 Hz. To get acceleration gravity units the x, y or z value needs to be divided by 336. The KSS indicates whether the belt buckle was closed (1) or open (0). Note, the order in which the data is returned depends on the programming language you use, e.g. in R the dataset is returned as a matrix of 4 columns and n rows, in python it is n- columns and 4 rows. In the HDF5file itself, h5ls will list it as a matrix with 4 rows and n columns.

Raw data is available here: https://s3.console.aws.amazon.com/s3/buckets/mueller-et-al-2019