Journal of Science and Medicine in Sport 22 (2019) 557–561

Contents lists available at ScienceDirect

Journal of Science and Medicine in Sport

journal homepage: www.elsevier.com/locate/jsams

Original research

Development of a gold-standard method for the identification of sedentary, light and moderate physical activities in older adults: Definitions for video annotation

Alan Kevin Bourke*, Espen Alexander F. Ihlen, Jorunn Lægdheim Helbostad

Department of Neuroscience and Movement Science, Faculty of Medicine and Health Science, Norwegian University of Science and Technology, Norway

ARTICLE INFO

Article history:
Received 25 June 2018
Received in revised form
24 September 2018
Accepted 6 November 2018
Available online 22 November 2018

Keywords:
Human activity recognition
Physical activity definitions
Video gold standard

ABSTRACT

Objectives: The development of a reliable method for the identification of sedentary, light and moderate physical activities in older adults. The method consists of a validated set of definitions for the identification of the initiation and termination of physical activities performed by older adult participants, video recorded during free-living and a laboratory setting.
Design: Inter-rater reliability assessment in a fully crossed design.
Methods: An iterative consensus process was used to define the initiation and termination of common activities of daily living. These definitions were then tested using videos recorded in two scenarios (1) by 9 raters who annotated a video recording, of a free-living protocol in a home environment, recorded in a first person view, using a body-worn camera and (2) by 7 raters who annotated a video recording, of older adults performing a semi-structured protocol in a living-lab environment, recorded in a third person view, using wall mounted cameras.
Results: Inter-rater reliability was excellent for all items, with Krippendorff's alpha and Fleiss' kappa all above 0.84 and a percentage of agreement above 88%. All ICC(C,1) inter-rater values for the activity quantity and duration were all above 0.9.
Conclusions: This set of physical activity initiation and termination definitions offers independent researchers a gold standard method to allow for the consistent annotation of high-frequency video footage (25fps), in both a free-living and laboratory setting. When synchronised with body-worn or ambient sensors, this annotation will allow for the development and validation of physical activity classification systems to a higher resolution than before.

Practical implications

- Validated definitions developed for the identification of the initiation and termination of sedentary, light and moderate physical activities, video recorded from older adults.
- Definitions suitable for the development of gold-standard data-sets of older-adult physical activities in a free-living and a laboratory setting.
- Developed definitions will improve human activity recognition algorithm development by providing a validated method to develop a gold-standard data-set.

1. Introduction

Accurate human activity recognition (HAR) in the older adult population, using ambient[1] or body-worn[2] sensors, is important for understanding the role of physical activity in health and physical function, for measuring the effect of interventions aimed at promoting independent living and reducing the incidence of falls.[3] The development of algorithms for HAR suffers from a lack of gold standard data-sets annotated to an adequately high detail. The majority of the frequency components of human body motion are contained in a range below 10 Hz[4] with 99% of body motion energy contained between 0 Hz and 15 Hz.[5] Since many existing data-sets are often labelled with a resolution of approximately 1 s (1 Hz), developed algorithms do not fully capture the details of human movement and thus are prone to excess error. Thus, in order to develop more accurate HAR algorithms it is necessary, according to

* Corresponding author.
E-mail address: alankevinbourke@gmail.com (A.K. Bourke).

https://doi.org/10.1016/j.jsams.2018.11.011

the Shannon–Nyquist Sampling Theorem,[6] to sample at twice the upper limit of the frequency of interest to fully capture the details of human physical activity.

Video recording technology offers benefits over existing methods of physical activity annotation, such as; self-report questionnaires, direct human observation labelled on paper,[7] using a portable device (e.g. touch screen tablet) or laptop.[8] All these methods suffer from a reported human cataloguing error of between one and three seconds,[8] since a delay between initiation and termination of activities and real-time cataloguing is inevitable.

Video validation of HAR systems have previously been performed, however the resolution of the annotation was between 1 s[9,10] and 10 s.[11]

With the abundance of body-worn high-frequency (>25fps) video recording technology (e.g. GoPro), the potential now exists to label free-living physical activity to a higher resolution than before. The video data can be synchronized with signals from body-worn[2] or ambient sensor[1] systems, allowing for the development of more accurate HAR systems and a more detailed validation of existing systems. Thus, a requirement exists for clear definitions of the initiation and termination of common physical activities recorded using high frame-rate video camera technology.

Development of definitions for the classification of physical activities in a real-world setting, using body-worn camera technology has been previously attempted by multiple studies. Doherty et al.[12] captured images approximately every 17 s. Dijkstra et al.[13] defined walking (including stair climbing), shuffling, sitting, standing and lying, and labelled physical activity to resolution of 0.1 s (10fps) using a video camera recording in the second person view. Fokkenrood et al.[14] outlined an annotation protocol for the identification of 5 types of activities of daily living and the transitions between them in a study using video camera technology annotated to a resolution of 0.1 s (10fps). The above-mentioned studies have thus ignored higher frequency movement, due to the frame rate employed, or used specially designed definitions for low frame rate video. Furthermore, Haché[15] defined the initiation and termination of changes-of-state of physical activities as part of a scripted routine, recorded using a digital video camera. However, these definitions were created as part of a routine where the preceding and subsequent activities were scripted and anticipated. Several studies have also used the technique of a combination of video camera technology and a visualisation of the raw signals from body worn inertial sensors to identify the initiation and termination of physical activities,[16–18] the sensor data were subsequently used for algorithm development. This is contradictory in its nature, as the annotation should be performed independently of the system being developed. By allowing the data that will be used to construct the data-driven prediction algorithm, to also be used in the construction of a ground-truth dataset, will thus ultimately contaminate this dataset since it will be influenced by the very data used to construct the algorithms.

Previous methods of video annotation have either allowed the sensor data to influence the annotation of the video recordings or have used a frame rate that ignores higher frequency movement. In summary, no previous study has been able to develop a robust set of definitions to identify the initiation and termination of human activities suitable as a gold-standard for the development of HAR algorithms.

Thus, the aim of this study is to develop a robust set of definitions for identification of the initiation and termination of human physical activities recorded using video camera technology, in both a first person view for activities in a free-living environment and a third person view for activities in a laboratory setting, which is suitable for annotation of video data recorded at a resolution that captures the details of human movement.

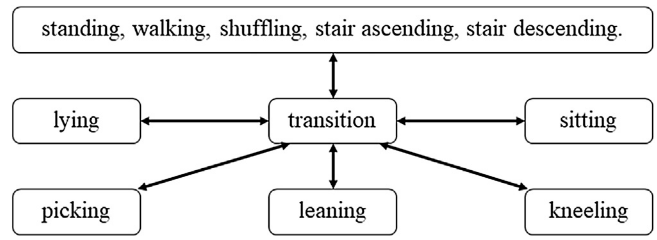

Fig. 1. A finite state machine design model for the relationship between upright activities, postures and transitions between physical activities. A person can change between an upright activity (standing, walking, shuffling, stair ascending, stair descending) without going through a postural transition (e.g. no defined postural transition exists here between standing and walking). However between an upright activity and any posture (sitting, lying, leaning, picking and kneeling) and between each posture, subjects perform a postural transition, for the purposes of the state machine model.

2. Method

The aim of this study was achieved in three stages. Stage (1); development of preliminary physical activity initiation and termination definitions. Stage (2); improvement of definitions through rater consensus during the annotation process, and finally Stage (3) inter-rater reliability testing of the developed definitions.

Stage 1: Development of the initiation and termination definitions. A list of 11 postures, transitions and activities, commonly performed in everyday life by older adults, was compiled.[19] A finite state machine design was then used to model the sequential pattern of human physical activity between these upright activities, postural transitions and postures, Fig. 1. This finite state machine design facilitates the development of HAR algorithms by modelling human movement in a manner commonly used in computational modelling.[20]

Based on the relationship between upright activities and postures, the transitions between the upright activities and sitting, lying, leaning, picking and kneeling, are required along with definitions for the individual upright activities and postures. Previous definitions of physical activity transitions have been developed for, the sit-stand transition,[21] sit-stand-sit,[22] sit-to-walk,[23] and lying-to-sit-to-stand.[24] These studies have developed these definitions for these specific transitions in isolation, as part of a clinical physical assessment for fundamental research. For the purposes of this study, definitions were developed that function within a framework where the sequence of individual activities and transitions is unlimited.

Preliminary definitions for the initiation and termination of each upright activity, postural transition and posture were created through incorporating existing dictionary[a] definitions that describe the primary biomechanical aspects of the different activities and postures along with existing definitions described in academic literature.[13–15] Through incorporating both types of definitions a robust set of preliminary definitions was created.

Stage 2: Refinement of the definitions. An iterative process was used to refine the preliminary definitions using 5 independent raters of a data-set consisting of 43.93 h of data recorded from 20 older adults (5 male/15 female, age from 68 to 90 years (76.4 ± 5.6 years), body mass from 56 to 93 kg (73.7 ± 11.4 kg) and height from 1.56 to 1.81 m (1.67 ± 0.072 m)) who were video recorded while performing two task-based activity protocols, a free-living (out-of-lab) protocol and a laboratory based (In-lab) protocol, incorporating a variety of activity sequences in addition to unscripted activity. Further details on the composition of the data-set are described

[a] Oxford English Dictorary.

Fig. 2. (A) The free-living unsupervised protocol recorded in a first person view (out-of-lab) using a body worn camera at the chest pointed downwards (GoPro HERO3+ Silver camera[b] attached using a harness, GoPro Chesty TM[c]) and (B) the laboratory based (in-lab) protocol video recorded in a third person view using wall-mounted cameras (Sony EVI-D80P[d]).

in Bourke et al.[19] The Regional Committee on Ethics in Medical Research in Central Norway approved the trial protocol and subjects provided written informed consent.[19] The In-lab protocol, video recorded in a third person view using wall mounted cameras, Fig. 2(B), is a semi-structured supervised protocol, performed in a living-lab environment (Total: 9.52 h). The out-of-lab protocol, video recorded in a first person view, Fig. 2(A), is a free-living unsupervised protocol, where participants are requested to incorporate defined task-based activities into their daily routine. This protocol is performed in the person's own home environment (Total: 34.41 h). During the annotation of both the in-lab and out-of-lab video data the definitions were refined and clarified in an iterative exhaustive process through annotator consensus to arrive at the final set of definitions. If an annotator was unsure about how to apply an activity definition, this was discussed among the research team and the definition was amended.

A description of the protocols and recording set-up is detailed in the protocol paper from the study.[19] Annotation of the video data was performed using the Anvil software (Version 5.1.13) annotation tool.[25]

Stage 3: Inter-rater reliability. The reliability of the definitions as a method for annotation of physical activity recorded in a third person view and first person view was then examined. The definitions were tested for inter-rater reliability in a fully crossed design using two randomly selected 20 min videos (20:00:09 (mm:ss:mss) 36001 frames, frame rate 25fps, frame size 640×360), one for each protocol type. A total of 9 and 7 raters independently annotated the out-of-lab video and in-lab video respectively. The sub-sets of the protocols performed by the subject during the out-of-lab and in-lab videos respectively, are presented in the Supplementary Appendix Tables 1 and 2.

Raters individually labelled each video. They were instructed not to allow any space between any element in the Activities track when labelling each activity using the definitions. This took place in the PC laboratory in the Movement Science Department in the Faculty of Neuroscience at St. Olav's Hospital.

The Inter-rater reliability was determined by calculating the intra class correlation coefficient (ICC(C,1)) for the duration and quantity of the activities, for both the in-lab and out-of-lab videos, using a two-way mixed effects, consistency, single measurement model.[26] Fleiss[27] considers ICC values above 0.75 may be taken to represent excellent. Portney & Watkins[28] consider ICC values above 0.90 as clinical measures and 0.75–0.90 as good. In addition, the comparative statistics of percentage of agreement[29] and Krippendorff's alpha[30] were also calculated for each rater pair and the arithmetic mean was taken to provide an overall index of agreement. This procedure is suitable for studies that have a fully crossed design.[29] Fleiss' kappa[31] was also calculated. Fleiss' kappa is a statistical measure used to assess the reliability of agreement between a fixed number of raters when assigning categorical ratings to a number of items or classifying items.

Fleiss[32] considers a kappa values over 0.75 as excellent.[25] Regarding alpha values, α is defined with two reliability scale points with 1.00 for perfect reliability and 0.00 for the absence of reliability.[33] Social scientists commonly rely on data with reliabilities $\alpha \geq 0.800$ and consider data with $0.800 > \alpha \geq 0.667$ to only draw tentative conclusions.[34] Detailed descriptions can be found at Ref.[33]

Inter-rater agreement was further examined using the method of Bland and Altman[35] by comparing the mean and difference of the quantity and duration of each activity for each rater pair for both the in-lab and out-of-lab videos.

These selected statistics provide a measure of how repeatable the definitions are. These comparative statistics were calculated using the Anvil 5.1.13 and Matlab R2016a software packages.

3. Results

The complete list of definitions developed through the consensus process in stage 2 and tested for inter-rater reliability in stage 3 is presented in the Supplementary appendix, Activity definitions Appendix Table 3.

For the out-of-lab video recorded in the first person view, the following activities were identified by all nine raters; leaning, lying, picking, shuffling, sitting, stair ascending, stair descending, standing, transition and walking. For the in-lab video recorded in the third person view, the following activities were identified by all seven raters; leaning, lying, picking, shuffling, sitting, standing, transition and walking.

The ICC(C,1) and 95% confidence intervals for the in-lab and out-of-lab video for both the activity quantity and duration are presented in Table 1. The ICC(C,1) values for the in-lab and out-of-lab are 0.92 and 0.94 for the activity quantity respectively, and 0.99 for the activity duration, thus representing excellent reliability. In addition, for both sets of statistics the p-values indicate that the observed agreement is not accidental.

[b] GoPro, Inc. 3000 Clearview Way, San Mateo, California, USA http://www.gopro.com/.
[c] GoPro, Inc. 3000 Clearview Way, San Mateo, California, USA http://www.gopro.com/.
[d] https://pro.sony/en_AL/products/ptz-network-cameras/evi-d80-d80p-pal-.

Table 1
In-lab and out-of-lab intra-class correlation coefficient—two-way mixed effect consistency single measurement, mean difference between rater pairs and the 95% confidence interval (CI) for activity quantity and duration, percentage of category agreement and Krippendorff's alpha.

	F-test		Bland–Altman				Percentage of agreement (%)			Krippendorff's alpha		
	ICC(C,1)	(95% CI)	Mean	95% CI		Within CI (%)	Mean	Max	Min	Mean	Max	Min
In-lab							88.2%	92.3%	82.1%	0.85	0.90	0.77
Activity quantity	0.92 (0.802–0.979)	$p < 0.0001$	−1.86	−39.66	35.94	90.5%						
Activity duration	0.99 (0.964–0.997)	$p < 0.0001$	0.01	−46.17	46.18	94.6%						
Out-of-lab							88.6%	93.3%	85.0%	0.84	0.91	0.80
Activity quantity	0.94 (0.877–0.981)	$p < 0.0001$	3.29	−14.85	21.45	90.9%						
Activity duration	0.99 (0.977–0.997)	$p < 0.0001$	0.00	−42.84	42.83	92.9%						

The Bland–Altman summary statistics for inter-rater agreement are presented in Table 1 along with in Bland–Altman plots, Supplementary Appendix Figure 1. Visual inspection revealed no systematic bias between raters when determining the quantity and duration of the different categories of physical activities, with >90% of activity quantity and >92% of the activity duration values are within the 95% confidence interval.

For both the in-lab and out-of-lab videos, excellent reliability coefficients were found for all items, with the mean Krippendorff's alpha above 0.84 and the percentage of agreement above 88%, Table 1. Fleiss' kappa values both for the in-lab and out-of-lab protocol were above 0.84, indicating an excellent level of agreement.

4. Discussion

A set of definitions to describe the initiation and termination of physical activities was developed and refined using an iterative consensus process. Inter-rater reliability data for the definitions was compiled through raters using the definitions for the identification and labelling (resolution 0.04s) of the participants' movements using video data recorded in two scenarios. The first, a 20 min video of an out-of-lab free-living protocol, recorded in a first person view, was annotated using 9 raters. The second, a 20 min video of a laboratory-based semi-structured protocol, recorded in a third person view, was annotated by 7 raters.

The statistical measures of inter-rater reliability produced excellent reliability coefficients. Krippendorff's alpha and Fleiss' kappa were all above 0.84 with the percentage of agreement above 88%. All ICC values for the activity quantity and duration were all above 0.90, thus demonstrating excellent agreement.

The Bland–Altman plots confirm that there is no systematic bias between raters when determining the quantity and duration of the different categories of physical activities, with more than 90% of values within the 95% confidence interval, Supplementary Appendix Figure 1.

Even though the inter-rater reliability results demonstrate excellent agreement some differences do exist. During the act of sitting, if there is movement of the trunk to allow repositioning (transition), this should be labelled as sitting-transition-sitting, this was sometimes ignored by some raters and contributed to the higher spread in transition quantity values observed, Fig. 2. Specifically, if the person momentarily comes to rest in a seated position and there is a change in trunk posture before repositioning on the chair seat, this should be labelled as, transition-sitting-transition-sitting, however some raters ignored the short sitting bout and labelled the activity, transition-sitting.

The object picking and leaning activities can be very short and often were no longer than one frame in duration, often signifying the change in direction of the trunk. The identification of this single frame event by each rater was not always identical. For the purposes of this study the term "picking" refers to the activity of picking or placing, an object on the floor. This is distinct from leaning as the body angle during picking from the ground can be different than leaning.

The definition of transition does allow some stepping in order for the person to position their body to transition between upright activities and postures and visa-versa. Thus, during some transitions from postures to upright activities, some raters extended the transition annotations into the walking/shuffling bout to include when the trunk became fully upright, however the trunk was in a stable posture before this and thus the transition had ended earlier. This contributed to the higher spread in transition duration values observed, Fig. 2. In our definition of transition, some steps can occur during the transition.

If turning occurred during, or at the end of the transition, this was sometimes labelled as shuffling before walking and visa-versa. Turning is included in the transition, if the person's trunk is in a stable upright posture and some turning occurs, this should be labelled as shuffling.

Some raters observed shorter transition durations than others during the transition to sitting, since some steps are permitted in the definition of transition. This stepping is to allow the subject to step into position to facilitate positioning of the body for sitting. The transition begins with a forward lean of the trunk or with a lowering of the centre of mass, but this was ignored by some raters. This is also evident in the Bland–Altman plots, where the largest variations are seen in the transition and standing activities in the in-lab data and standing, shuffling, walking and transition in the out-of-lab data.

Sometimes standing was confused as leaning. Leaning was included as a category to describe the posture when a person has lowered their trunk but are still supporting their body-weight through their feet (e.g. when interacting with an object placed on a low shelf or table). The category leaning was intended to describe the difference between when a person is in a comfortable standing posture and when they are not, bending at the hips for instance.

A momentary sitting posture between walking/shuffling and lying was identified and labelled by some raters but not by others. If there is a pause in the person's movement during the act lying, this can be labelled as transition–sitting–transition–lying, otherwise it should be transition–lying.

The activity that was inconsistently classified the most was the shuffling task. Certain raters interpreted a task involving moving objects from one table to another as walking–transition–leaning–transition–walking whereas others identified a combination of walking–shuffling-standing. If the amount of stepping is shorter than 1 stride, shuffling should be annotated. If the person is in an upright posture without feet movement, standing should be annotated.

The aim of this study was to develop definitions for the development of a dataset of normal activities of daily living, thus vigorous activities and exercise were excluded from this process. Another limitation is that the upper-limit of the frame rate of the camera was 25 frames/second. Implying that human movement between 0 Hz and 12.5 Hz can be detected according to the Shannon–Nyquist sampling theorem and thus the majority of human movement detail,[4] which is likely adequate for the activities performed here, being of moderate intensity at most and recorded from an older adult population. Future work in this area should aim to record at a frame-rate of 30fps or above, to fully capture the range of 0 Hz–15 Hz suggested earlier.[5] This would also compliment the inclusion of vigorous activities and exercise as advocated for here, and performed by subject cohorts of different age profiles.

The resources required to perform the complete video annotation, by the 5 independent raters, was 3.4 person months. This includes the video labelling, peripheral video file preparation and all consensus discussions. It is estimate for video labelling alone that it takes approximately 45 min–1.5 h, for each 20 min in-lab video and 1.5 h–2 h, for each 20 min out-of-lab video.

Future work could attempt to provide more detail to the definitions for the initiation and termination of transitions, addressing further the involvement of steps, as well as more detail to the definitions for walking and shuffling.

The work presented here complements the efforts being performed in the ARDUOUS project,[36] the AlPHABET project,[37] Sensor Methods Collaboratory (National Institutes of Health),[38] the SBRN[39] and the SIT project[40] which aim to develop common taxonomies, definitions and annotation procedures for cataloguing of human physical behaviour in daily life through international consensus and collaboration.

5. Conclusion

This study presents the development of a set of definitions for the identification of the initiation and termination of physical activities performed by older adult participants, recorded using video camera technology in first and third person views, during a free-living situation and laboratory setting respectively. Overall the statistical level of agreement was excellent. We have thus produced a validated set of definitions for the identification of sedentary, light and moderate physical activities in older adults using annotation of video recordings. These definitions facilitate high frame-rate (25fps) labelling of video data, which can be synchronised with ambient and body worn sensor data, for the development of HAR algorithms for improved physical activity classification and validation. This validated set of definitions also allows for the creation of a larger data-set through collaboration from independent research groups.

Acknowledgments

This research was funded by The Research Council of Norway (FRIMEDBIO, Contract No. 230435) and the European Union Seventh Framework Programme (FP7/2007-2013) FARSEEING project; grant agreement No. 288940.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jsams.2018.11.011.

References

1. Ranasinghe S, Al Machot F, Mayr HC. A review on applications of activity recognition systems with regard to performance and evaluation. *Int J Distrib Sens Netw* 2016; 12.
2. Godfrey A. Wearables for independent living in older adults: gait and falls. *Maturitas* 2017; 100:16–26.
3. Lara J, Godfrey A, Evans E et al. Towards measurement of the Healthy Ageing Phenotype in lifestyle-based intervention studies. *Maturitas* 2013; 76:189–199.
4. Zeng H, Zhao Y. Sensing movement: microsensors for body motion measurement. *Sensors* 2011; 11:638–660.
5. Bouten CV, Sauren AA, Verduin M et al. Effects of placement and orientation of body-fixed accelerometers on the assessment of energy expenditure during walking. *Med Biol Eng Comput* 1997; 35:50–56.
6. Shannon CE. Communication in the presence of noise. *Proc IRE* 1949; 37:10–21.
7. Lyons GM, Culhane KM, Hilton D et al. A description of an accelerometer-based mobility monitoring technique. *Med Eng Phys* 2005; 27:497–504.
8. Paraschiv-Ionescu A, Buchser EE, Rutschmann B et al. Ambulatory system for the quantitative and qualitative analysis of gait and posture in chronic pain patients treated with spinal cord stimulation. *Gait Posture* 2004; 20:113–125.
9. Taylor LM, Klenk J, Maney AJ et al. Validation of a body-worn accelerometer to measure activity patterns in octogenarians. *Arch Phys Med Rehabil* 2014; 95:930–934.
10. Capela NA, Lemaire ED, Baddour N et al. Evaluation of a smartphone human activity recognition application with able-bodied and stroke participants. *J Neuroeng Rehabil* 2016; 13:1–10.
11. Aminian K, Robert P, Buchser EE et al. Physical activity monitoring based on accelerometry: validation and comparison with video observation. *Med Biol Eng Comput* 1999; 37:304–308.
12. Doherty AR, Kelly P, Kerr J et al. Using wearable cameras to categorise type and context of accelerometer-identified episodes of physical activity. *Int J Behav Nutr Phys Act* 2013; 10:1–11.
13. Dijkstra B, Kamsma YP, Zijlstra W. Detection of gait and postures using a miniaturized triaxial accelerometer-based system: accuracy in patients with mild to moderate Parkinson's disease. *Arch Phys Med Rehabil* 2010; 91:1272–1277.
14. Fokkenrood HJP, Verhofstad N, Van Den Houten MML et al. Physical activity monitoring in patients with peripheral arterial disease: validation of an activity monitor. *Eur J Vasc Endovasc Surg* 2014; 48:194–200.
15. Haché G. *Development of a Wearable Mobility Monitoring System*, University of Ottawa, 2010.
16. Massé F, Gonzenbach R, Paraschiv-ionescu A et al. Wearable barometric pressure sensor to improve postural transition recognition of mobility-impaired stroke patients. *J Neuroeng Rehabil* 2016; 12.
17. Reyes-ortiz J, Oneto L, Samà A et al. Transition-aware human activity recognition using smartphones. *Neurocomputing* 2016; 171:754–767.
18. Anguita D, Ghio A, Oneto L et al. A public domain dataset for human activity recognition using smartphones. *ESANN 2013 Proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning* 2013:24–26.
19. Bourke AK, Ihlen EAF, Bergquist R et al. A physical activity reference data-set recorded from older adults using body-worn inertial sensors and video technology—the ADAPT study data-set. *Sensors (Switzerland)* 2017; 17.
20. Fernández-Caballero A, Castillo JC, Rodríguez-Sánchez JM. Human activity monitoring by local and global finite state machines. *Expert Syst Appl* 2012; 39:6982–6993.
21. Schenkman M, Berger RA, Riley PO et al. Whole-body movements during rising to standing from sitting. *Phys Ther* 1990; 70:638–648, discussion 648–651.
22. Kerr KM, White JA, Barr DA et al. Analysis of the sit-stand-sit movement cycle in normal subjects. *Clin Biomech* 1997; 12:236–245.
23. Kerr A, Durward B, Kerr KM. Defining phases for the sit-to-walk movement. *Clin Biomech* 2004; 19:385–390.
24. Bagala F, Klenk J, Cappello A et al. Quantitative description of the lie-to-sit-to-stand-to-walk transfer by a single body-fixed sensor. *IEEE Trans Neural Syst Rehabil Eng* 2013; 21:624–633.
25. Kipp M. ANVIL: a universal video research tool, in *Handbook of Corpus Phonology*, Durand J, Gut U, Kristofferson G, editors, 2014, p. 420–436.
26. Mcgraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1996; 1:30–46.
27. Fleiss JL. *The Design and Analysis of Clinical Experiments*, John Wiley & Sons, Inc., 1999.
28. Portney LG, Watkins MP. *Foundations of Clinical Research: Application to Practice*, 3rd ed. Pearson Prentice Hall, 2009.
29. Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. *Tutor Quant Methods Psychol* 2012; 8:23–34.
30. Krippendorff K. *Content Analysis: An Introduction to Its Methodology*, Sage, 2013. p. 221–250.
31. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull* 1971; 76:378–382.
32. Fleiss JL. *Statistical Methods for Rates and Proportions*, 2nd Edition New York, John Wiley, 1981. p. 38–46.
33. Hayes AF, Krippendorff K. Answering the call for a standard reliability measure for coding data. *Commun Methods Meas* 2007; 1:77–89.
34. Krippendorff K. Standards for data reliability, In: *Content Analysis: An Introduction to its Methodology*. Sage, 2004. p. 241–243.
35. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 30:7–310.
36. Yordanova K, Paiement A, Schröder M et al. Challenges in annotation of useR data for UbiquitOUs systems: results from the 1st ARDUOUS workshop, *IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 2018.
37. Chastin S, Helbostad JL, Tremblay MS et al. *AlPHABET: Development of a Physical Behaviour Taxonomy with an International Open Consensus*, 2017. Available at: osf.io/2wuv9. Accessed 6 September 2018.
38. Troiano R.*Šensor Methods CollaboratoryÎnternational Society for the Measurement of Physical Behaviour, Pre-Conference-Event*, 2017.
39. Sedentary Behaviour Research Network. 2017. Available at: www.sedentarybehaviour.org. Accessed 20 September 2018.
40. Chastin SFM, Schwarz U, Skelton DA. Development of a consensus taxonomy of sedentary behaviors (SIT): report of Delphi round 1. *PLoS One* 2013; 8:1–16.