Concurrent measurement of GPS and event-based physical activity data: a methodological framework for integration.

Date of Submission: 14th February 2020

Date of Resubmission: 31st August 2020

Iveson AMJ[1], Granat MH[2], Ellis BM[1], Dall PM[1]

[1] School of Health and Life Sciences, Glasgow Caledonian University, Glasgow, UK

[2] School of Health Sciences, Salford University, Salford, UK

Corresponding author:

Philippa Dall

Senior Research Fellow

School of Health & Life Sciences

Glasgow Caledonian University

Cowcaddens Road

Glasgow

G4 0BA

philippa.dall@gcu.ac.uk

Acknowledgements

The HSC analysis software was written by Dr Philippa Dall and Professor Malcolm Granat. Software to integrate processed GPS and activPAL data, and to derive outcomes using the framework, was written by Dr Philippa Dall.

Conflict of Interest

Malcolm Granat is a co-inventor of the activPAL physical activity monitor and a director of PAL Technologies Ltd. The remaining authors declare no competing interests.

Funding

This work was part of the PhD studies of Dr Anna Iveson which was funded by Glasgow Caledonian University.

Abstract

Objective: GPS data can add context to physical activity data, and have previously been integrated with epoch-based physical activity data. The current study aimed to develop a framework for integrating GPS data and event-based physical activity data (suitable for assessing patterns of behaviour). Methods: A convenience dataset of concurrent GPS (AMOD) and physical activity (activPAL) data were collected from 69 adults. GPS data was (semi-) regularly sampled every 5 seconds. The physical activity data output was presented as walking events, which are continuous periods of walking with a time-stamped start time and duration (to nearest 0.1s). GPS outcome measures and the potential correspondence of their timing with walking events were identified and a framework was developed describing data integration for each combination of GPS outcome and walking event correspondence. Results: GPS outcome measures were categorised as those deriving from a single GPS point (e.g. location), or from the difference between successive GPS points (e.g. distance), and could be categorical, scale or rate outcomes. Walking events were categorised as having zero (13% of walking events, 3% of walking duration), or one or more (52% of walking events, 75% of walking duration) GPS points occurring during the event. Additionally, some walking events did not have GPS points suitably close to allow calculation of outcome measures (31% of walking events, 22% of walking duration). The framework required different integration approaches for each GPS outcome type, and walking events containing zero or more than one GPS points.

Key words: physical activity, GPS, accelerometer, event-based data, methods

Introduction

Lack of physical activity is a key modifiable risk factor for many aspects of health, including mortality, cardiovascular disease, obesity and metabolic disorders (Kyu et al, 2017; Lee et al., 2012), and is important in maintaining health, function and quality of life across a range of long-term conditions (Garg et al., 2006; Saunders et al., 2016). Appropriate measurement of physical activity for surveillance, and to assess the effect of policy changes and interventions is important for individuals and for the population as a whole (Chastin et al., 2018; Strain, Milton, Dall, Standage, & Mutrie, 2019). Recently, body-worn sensors for the objective measurement of physical activity have become mainstream. However, the context in which physical activity occurs is an important aspect of behaviour which is usually missing from such assessment (Jankowska, Schipperijn, & Kerr J, 2015). The context of physical activity can be used to identify areas for individual improvement and for tailoring of interventions, and can provide insight about the social and cultural aspects of when and where physical activity is undertaken (James et al., 2016; Jankowska et al., 2015).

Physical activity monitors can differ in the manner in which they interpret acceleration data, and present it as outcome measures. One commonly used method (e.g. ActiGraph) is to aggregate the amount of acceleration accumulated during a specified period of time (epoch) into a cumulative value, usually reported in terms of a proprietary value (counts). Calibration studies relating values of counts per epoch to values of objectively measured energy expenditure are used to convert this information into time spent in categories of physical activity intensity (Matthews, 2005). Whilst this can be a good way of representing cumulative values of activity, there are a number of limitations with assessing some aspects of physical behaviour. Physical activity is often undertaken in short durations (as measured by objective monitors, Chastin et al., 2009), and commonly used epoch durations (e.g. 1 minute) can aggregate several short, high-intensity periods of physical activity into one epoch, which is then reported with a lower average intensity (Kim, Beets, Pate, & Blair, 2013). One alternative approach is to assess physical activity as a series of events, continuous periods of a single type of activity (Granat, 2012). This event-based approach can address the limitations of an epoch-based approach, as rates (e.g. cadence) and intensity do not get spread across inappropriately applied epochs (e.g. Dall, McCrorie, Granat, & Stansfield, 2013). In addition, use of an event-based approach to analysis can facilitate insight into the patterns in which physical activity is undertaken. For example, in individuals with Intermittent Claudication who need to stop walking frequently due to pain, patterns of broken-up walking can be identified (Clarke, Holdsworth, Ryan, & Granat, 2013).

One option for adding context to physical activity data is the use of global positioning system (GPS) devices (Kerr, Duncan, & Schipperijn, 2011), which use satellite tracking to assess the location of individuals at regularly sampled intervals. These data therefore have the potential to provide information about the physical terrain through which someone is moving, and provide insight into the location in which activity occurs (Jankowska et al., 2015). Additionally, GPS provides the potential to expand physical activity-related outcome measures, for example reporting on distance travelled (Kerr et al., 2011). GPS devices have been successfully used to provide context to objectively measured physical activity data, including identifying and evaluating use of local parks (Evenson, Wen, Hillier, & Cohen, 2013), the location of children's physical activity (Carlson et al., 2016) and the contribution of commuting to meeting PA guidelines (Rafferty, Dolan, & Granat, 2016). Previously in these studies, physical activity data have been processed using an epoch-based approach, which facilitates integration of physical activity and GPS data, as both are reported at regular sampling intervals. By harmonising the sampling frequency of the GPS devices with the activity monitor epochs, data can be integrated for each matched sample/epoch using a one-to-one correspondence (Hurvitz, Moudon, Kang, Saelens, & Duncan, 2014). Integration of event-based physical activity data (which is reported for a discreet period of time commencing at irregular intervals) with GPS data (sampled at regular time points) is not straightforward, and requires careful examination of the correspondence between walking events and the GPS data. For example, longer walking events are likely to take place across several GPS data samples, and thus a situation arises where information from multiple GPS points is associated with a single walking event (a many-to-one correspondence), requiring data aggregation from multiple sources. In contrast, short walking events might take place entirely between GPS data samples (a none-to-one correspondence), meaning that inferences about location must be made about that walking event without directly corresponding GPS data.

The integration of such data has not been previously reported, but is required to enable any physical activity study using an event-based approach to effectively add the context from GPS data. The aim of the current study was to develop a methodological framework for integration of GPS and event-based physical activity data, to enable the assessment of the context of physical activity when an event-based approach is appropriate.

Methods

Study Design

This article describes the development of a methodological framework to integrate GPS data with event-based physical activity data. Framework development was informed by, and then tested on, a convenience dataset consisting of concurrently measured GPS and physical activity event-based data. The integration framework uses GPS derived outcome measures to add location information to each physical activity event. The correspondence of the timing of GPS data points with the physical activity events informs the theoretical basis of the framework. For example, whether the timing of the walking event corresponds with none, one or many GPS points will determine from which points, and how, GPS derived data is added to the walking event. In the current study, walking behaviour has been assumed to be of primary interest, and thus data processing and framework development concentrated only on walking events. Following development, the integrity of the integration framework was tested using basic performance metrics derived from the convenience dataset.

Participants

Two groups of individuals took part in the current study. The dataset was collected as part of a study investigating the habitual physical activity of individuals with Intermittent Claudication which used a matched pairs design, which allowed the framework to be developed and tested on individuals with a broad range of walking activity. Individuals with mild or moderate Intermittent Claudication were recruited from Lanarkshire Community Claudication Clinics. Participants were eligible if they were adults with a medical diagnosis of Intermittent Claudication which was the principle condition that limited their walking (in practice Fontaine's stage II, Norgren et al., 2007). Control participants from the general population were recruited matched (by sex, age ±5 years, and home location within 5 miles) to the individuals with Intermittent Claudication. Control participants were excluded if they self-reported symptoms of Intermittent Claudication or discomfort in the legs when walking. All participants were excluded if they had a chronic condition that had a negative impact on their walking ability (e.g. unstable angina pectoris, respiratory or neuromuscular disease, or history of stroke). Ethical approval was obtained from NHS West of Scotland Research Ethics Committee, and all participants provided written informed consent.

Data Collection Protocol

Activity monitor.

Participants wore a (uniaxial) activPAL activity monitor (PAL Technologies, Glasgow, UK) continuously for seven days. The activPAL is a small, lightweight uniaxial accelerometer, which is valid for categorising posture and activity (sit/lie, stand, walk) with 96% accuracy in adults (Grant, Ryan, Tigbe, & Granat, 2006). It also measures step count with 99% accuracy in older adults (Grant, Dall, Mitchell, & Granat, 2008). Monitors were waterproofed by wrapping them in a nitrile sleeve (PAL Technologies, Glasgow, UK) and a waterproof medical dressing (Opsite Flexifix, Smith & Nephew, Netherlands). On day 0, the activPAL was attached half-way up the mid-line of the front of the thigh of the dominant leg of the participant by a researcher. The monitor was attached using a double-sided hypoallergenic sticky pad (PAL stickies, PAL Technologies) and covered by a waterproof dressing (Opsite Flexifix). Participants were asked to wear the monitor continuously, including overnight and during water-based activities on days 1 to 7. Monitors were removed by the participant on day 8.

GPS device

On day 0, participants were provided with a GPS device (AGL3080, AMOD Technology, Taipei City, Taiwan), a battery charger and spare AAA batteries. The device takes between 1s and 42s to start collecting data after being switched on (hot and cold time to first satellite fix), and was set up to sample GPS data every 5 seconds (from a choice of 1, 5 or 10 seconds) which allowed seven days of data to be stored in the memory. The GPS device has a positional accuracy for latitude and longitude of 10m (Goel, Gani, Guttikunda, Wilson, & Tiwari, 2015), although this will be dependent on environmental factors (Wing, Eklund, & Kellogg, 2005). Preliminary testing prior to this study aggregating data from ten walks over three days round a UK city, indicated that mean difference for a static position (GPS device placed on Ordnance Survey Trigonometric point T072) was 1m for latitude and longitude, and for stops during a walk (six locations with repeat position estimated by user) was within 5m for both latitude and longitude. Participants were asked to have the GPS device switched on and to carry it with them at all times during the seven-day data collection period, even when at home. The GPS device did not have a belt-clip, and participants were asked to carry it in a pocket or bag. It was particularly emphasised to participants that they should carry the GPS device with them each time they left the house. The battery life was not sufficient to last for 7 days, and participants were asked to switch off the unit each evening, replace the AAA batteries, and then immediately switch the GPS device back on. The set of batteries not in the GPS device were then charged overnight.

Both the activPAL monitor and the GPS device were returned to the researcher by post.

Data Processing

The (uniaxial) activPAL samples acceleration at 10Hz. Data were downloaded using proprietary software (PAL Technologies, version 5.9.1.1), which categorises the data into categories of sit/lie, standing and walking. Data were exported as events, with successive continuous periods of each category labelled with the start time, duration, and type of event (sit; stand; stride). Custom software in excel (HSC analysis program) was used to process data from midnight to midnight for each day of measurement, and to aggregate strides into continuous walking events. Each walking event had the associated data of start time, duration and number of steps taken. From this, the average event cadence was calculated (total number of steps/duration).

GPS data were downloaded from the AMOD monitor in tab-delimited format. Each GPS data point recorded contained the following information: time; latitude; longitude; number of satellites in view; dilution of precision values (which characterise the potential magnitude of error of fixing location due to the geometric distribution of satellites in view in relation to the GPS device; values are reported for overall position, and the horizontal and vertical components of position); signal to noise ratio. Data were recorded in separate sessions each time the monitor was switched on, and data was separated or aggregated, as required, into 24 hour periods to match the activPAL data. Latitude and longitude (co-ordinate reference system ETSR89 Geodectic) were converted into Eastings and Northings (OSGB 1936/British National Grid) using InQuest software (v7.0.0, Ordnance Survey, UK). Height above sea level was obtained from the value on a 5m digital terrain map (GetMapping LLP, UK, sourced from MIMAS, University of Manchester, UK) of the GPS location (Eastings and Northings) using the surface spot tool in ArcMap (v9.3, ESRI, USA).

GPS data is known to be subject to errors, due to satellite positioning, atmospheric conditions, topography, or reflection off buildings in an urban environment, and needs to be cleaned prior to use (Kerr et al., 2011). The PALMS framework (https://uscd-palm-project-wikispaces.com) provides recommended cleaning parameters for a range of GPS monitors (Jankowska et al., 2015), but the AMOD GPS device used in the current study was not included in the framework. Therefore, thresholds for cleaning the GPS data were derived empirically comparing pilot GPS data (not reported) to their locations on known walks in geographic information systems (GIS) viewing software (ArcMap v9.3). For the current study, which concentrated on walking behaviour and not transport, the GPS data points were removed if: speed (relative to the previous point) was >8m/s; any dilution of precision measure (horizontal, vertical or position) was >10; or the number of satellites in view was 0.

The GPS device takes its time stamp from the satellites (updated regularly), and always records the time in Greenwich Mean Time (GMT). The activPAL takes its time stamp from the computer which it is connected to for programming. In the UK (where the data was collected), this may be GMT or British Summer Time (BST; GMT+ 1 hour), depending on the time at which the monitor is programmed. The activPAL does not adjust for changes between GMT and BST that occur during the measurement period, so this was done in post-processing, by adding (converting GMT to BST) or subtracting (converting BST to GMT) one hour to the event start time after the change occurred (i.e. missing or doubling an hour, as appropriate). The time stamp of both monitors was then converted in to a common format, equivalent to the participant's perceived time (GMT or BST).

Framework Development

Guiding principles and philosophy.

Event-based physical activity data has three basic properties: start date/time; duration; and type of event. To these basic properties, other aspects can be added, for example, walking events have the added property of the number of steps taken during the event. Within the framework for integration, the philosophy followed was to add properties and outcomes derived from the GPS data to the walking events measured by the activPAL monitor.

The main difficulty was integrating data collected with two different sampling frames. GPS data is collected as a single measurement point, collected at (semi-) regular sampling intervals, in the current study approximately every 5 seconds, although some outcome measures will be derived from the difference between successive measurement points (e.g. distance travelled). Event-based walking data, in contrast, is presented as a variable duration in time, which can start and end at irregular time-points (in the current study to a resolution of 0.1 seconds). For example, someone walking to the local shops in the morning might start with a short walking event (starting at 09:15:05 with a duration of 4 seconds) as they move from the chair to the front door. After a pause to put on their shoes, they walk to the shops stopping once to cross a busy road giving two walking events (starting at 09:18:23 with a duration of 15 minutes and 45 seconds, and starting at 09:36:08 with a duration of 1 minute 25 seconds).

Framework development process.

Developing the framework for integration of event-based physical activity data with GPS data was undertaken in three distinct stages: (i) categorisation of derived GPS outcomes; (ii) identification of types of correspondence between walking events and GPS data; (iii) and developing the integration framework. Firstly, the range and type of outcome measures which could be derived from the GPS device, in relation to how they interacted with each GPS measurement point, were categorised. For example, distance is a scale variable that can be derived from the change in position between two GPS data points. Secondly, the different correspondences of the duration of the walking event with GPS measurement points were characterised. For example, whether several GPS data points occurred during a walking event, or whether the walking event occurred entirely between GPS data points. This was undertaken as both a theoretical exercise (identifying all possible combinations of event duration and GPS points) and as an empirical exercise (identifying from the data collected how likely each theoretical situation was to occur), allowing a rational decision to be made as to which situations to incorporate within the final framework. Finally, a framework was developed indicating how each category of GPS outcome data should be processed to derive an outcome property for each type of correspondence with a walking event. For example, providing a method of how to calculate a value for total distance walked for a walking event that occurred between two GPS data points.

Assessment of Framework Integrity

The event-based physical activity data and GPS data from all participants were processed according to the framework developed, and assessed using the following metrics of integrity. 1) Missing data. The percentage of walking events that was not covered by GPS data was reported. Missing data could occur for walking events undertaken when the GPS device was switched off (as this was independent of the continuous operation of the activity monitor), or if walking events occurred between two GPS points which were far apart in time (for example if the GPS signal was temporarily lost through being indoors or under trees). 2) Free-living correspondence of walking events with respect to GPS points. The number of instances of each type of correspondence of walking events with the GPS points from the framework were reported, to assess likelihood of each situation occurring during habitual free-living activity. 3) Sensitivity analysis of thresholds for inclusion of external GPS points. The purpose of this was to explore how inclusion of walking events in analysis changed if the thresholds set in this study were relaxed or removed. 4) An exploration of outliers. The purpose of this was to assess whether GPS data/walking event correspondence combinations presented systematic differences in their implementation, and concentrated on outcomes (e.g.

speed) that were considered to be prone to potential measurement errors within the framework.

Results

Participants

Data was collected from 35 individuals with Intermittent Claudication, and from 34 matched control participants (for one individual with Intermittent Claudication a matched control was not recruited within the study timescale). Diaries were used to identify any non-wear periods of the activPAL, and additional visual inspection of the data was performed to remove days with obvious non-wear. ActivPAL data was considered to be continuous for 24 hours, unless non-wear was reported or identified through visual inspection of the data. There was no lower limit set on duration of GPS for a valid day of data. Data from a participant was included in analysis if there were at least four days with both activPAL and GPS data for assessment. This threshold was selected pragmatically to provide a reasonable snapshot of usual behaviour. Data was included in analysis if both participants in a matched pair met the minimum criteria for inclusion. In total, data from 56 individuals (28 pairs of participants) were included in analysis. All participants had seven days of activPAL data. Thirty-eight participants had seven days of GPS data (mean 6.5±0.8), covering a mean of 10.5±2.5 hours/day. Participants were mostly men (64%) and were aged 67±8 (54-89) years. The mean BMI (26.9±4.6 kg.m^{-2}) of the participants was categorised as overweight, ranging from underweight (18 kg.m^{-2}) to morbidly obese (43 kg.m^{-2}). Participants were all successfully matched for sex and age. There was no difference in BMI between participants with Intermittent Claudication and controls (paired t-test; p=0.86). The participants with Intermittent Claudication (n=28) had mostly moderate disease severity (mean ankle-brachial pressure index 0.71±0.15; mean Lothian Oximetry Index 0.78±0.19), with a disease duration of between 0.3 and 20 years.

Framework Development Stage 1: Categorisation of GPS Derived Outcome Measures

Derived outcomes from the GPS data (table 1) can be first categorised as (i) outcomes deriving from a single GPS point (for example, the distance from home), and (ii) outcomes deriving from the differential values between successive GPS points (for example, the speed of movement between two GPS data points). Further division can then be made based on whether the derived variables were categorical (for example, whether the GPS point was indoors or outdoors), scale (for example, distance) or rate (for example, speed) outcomes.

Framework Development Stage 2: Categorisation of Correspondence of Event-based Physical Activity (Walking Events) with GPS Data Points

A walking event takes place over a variable duration, in the current study ranging from 0.5 seconds to 60 minutes. For any walking event, zero, one or more than one GPS data points could occur within the event duration (figure 1). The first categorisation of the correspondence of walking events with GPS data points was to distinguish those events in which zero GPS data points occurred during the walking event, from those where at least one data point occurred within the duration. Walking events with zero GPS points effectively occurred entirely between two successive GPS data points (e.g. walking event B which occurs between GPS data points 2 and 3, figure 1). Walking events occurring entirely between regularly sampled GPS points (in the current study every 5 seconds) would necessarily be of short duration, and the GPS points immediately external to the walking event would contain information relevant that walking event. However, any extended loss of GPS signal, for example when walking under dense tree cover, could lead to two GPS points being separated by a considerable duration. A decision therefore has to be made whether the GPS points external to the walking event are relevant for inclusion in any calculation of outcomes for that walking event. Rules for inclusion were formulated based on two criteria, either (i) the duration between the GPS points surrounding the walking event, or (ii) the absolute duration between the end of the walking event and the next external GPS point. GPS points only had to meet one criteria to be included. Firstly, a maximum duration between the two GPS points outside the walking event was considered, and any walking events occurring entirely between pairs of GPS points less than this duration apart were considered suitable for inclusion. In the current study, GPS data surrounding a walking event were included if the duration between GPS points was <=60s. For example, in Figure 1, walking event B occurs entirely between GPS points 2 and 3 which are less than 60 seconds apart. Thus GPS points 2 and 3 are included in the calculation of outcome measures for walking event B. Secondly, for walking events between GPS points which were further apart, the absolute duration of time between the end of the walking event and the nearest external GPS point was considered. This was to ensure that a GPS data point occurred within a reasonable time of the end of the walking event. In the current study, GPS data were included if the time between the end of the walking event and the nearest external GPS point was <=30s.

Except in the infrequent circumstance that the duration of the walking event (to the nearest 0.1s) is both an exact factor of the sample rate and starts exactly when the GPS is sampled, the start and end of the walking event will not coincide with a GPS point. This is true regardless of how long the duration of the walking event, or how many GPS points occur during the walking event. The time between the GPS point prior to the walking event and the start of that event will be variable (similarly for the end of the walking event; Figure 1), and a decision has to be made as to whether to include information from these GPS data points when calculating GPS outcomes for that walking event. For example, in Figure 1, a decision needs to be made for walking event C as to whether GPS point 2 or GPS point 5 should be included when calculating outcome measures for that walking event. A simple decision could be made to only include data from GPS points occurring within the walking event, however many walking events were short (35% of events lasted <5 seconds), and thus there was potential to lose much valuable GPS information using that approach. Rules for inclusion were formulated based on two items; (i) the proportion of time between GPS data points that is covered by the walking event, and (ii) the absolute time between the end of the walking event and the next GPS data point. Firstly, the proportion of the time between internal GPS data points (occurring within the duration of the walking event, for example GPS points 3 and 4 within walking event C in Figure 1) and external GPS data points (occurring outside the duration of the walking event, for example GPS points 2 and 5 outside walking event C) that was covered by the walking event was considered. This was to ensure that most of the time between GPS points was relevant to the walking event. In the current study, GPS data was included within the calculation of GPS outcomes for a walking event if the walking event covered over 50% of the duration between internal and external GPS points. For example, in Figure 1, walking event C covers most (> 50%) of the time between GPS points 4 and 5, and thus meets the criteria for inclusion. Secondly, the absolute duration of time between the end of the walking event and the nearest external GPS point was considered. This was to ensure that a GPS data point occurred within a reasonable time of the end of the walking event. In the current study, GPS data were included if the time between the end of the walking event and the nearest external GPS point was <=30s. For example, in Figure 1, the time between the end of walking event C and GPS point 5 was 3s

(<30s) and thus meets the criteria for inclusion. The GPS point had to meet both criteria for inclusion. In Figure 1, although walking event A covers most (>50%) of the time between GPS points 1 and 2, the time between the end of walking point A and GPS point 1 was more than 30 seconds, and thus GPS point 1 is not included within walking event A. This calculation was performed independently, and a separate decision made, for the GPS points just before and just after the ends of the walking event.

Framework Development Stage 3: Framework for Integrating GPS Data with Event-based Physical Activity Data

The framework for integrating event-based physical activity data and GPS data is shown in Table 2. Each of the types of GPS derived outcome measure required a different method to calculate summary data for adding to walking events. Different methods of calculating GPS outcome measures were also required depending on the combination of the number of GPS points internal to the walking event and the number of external GPS points which met criteria for inclusion. There were some combinations of correspondence of internal and external GPS points and walking event for which outcomes could not be calculated, which are marked as missing data in Table 2.

For data based on successive GPS points, scale data was conceptualised as occurring linearly (i.e. at a constant rate) between the two points. For example, if a distance of 5m was travelled between successive GPS points 5 seconds apart, it was assumed this was undertaken at a constant rate (i.e. a speed of 1.0 m/s), and the distance travelled in a proportion of the time (e.g. half the time= 2.5s) was the same portion of the distance travelled (e.g. half the distance = 2.5m). Although this is unlikely to have been the case in reality, no data had been recorded between GPS points to allow a more accurate interpretation. At the usual sample rate of 5s, adults might walk approximately 6.8m (at a usual walking speed of 1.35 m/s, Bohannon, & Andres, 2011) to 9.8m (at a fast walking speed of 1.95 m/s, Lythgo, Wilson, & Galea, 2011) in that time, and a linear approximation was considered adequate.

Calculation of outcome measures to tag walking events was conducted assuming that the value of the outcome measure could be scaled based on the proportion of the duration of the walking event that corresponded with the relevant GPS points. For categorical outcomes this resulted in the assumption that categorical codes based on a single point covered 50% of the time between the GPS points before and after the GPS point. For example, in figure 2, the code of outdoors for GPS point 3 was assumed to extend from the time midway between GPS points 2 and 3 to the time midway between GPS points 3 and 4. For categorical codes based on two successive GPS points, it was assumed that the code applied to the entire duration between those points. For example, if the gradient between two GPS points was categorised as uphill, it was assumed that all walking occurring between those GPS points was uphill. For scale outcomes, this resulted in the assumption that the scale outcome could be assigned proportionally based on the duration of the walking event (figures 3 and 4). For example, in figure 4, if the duration of walking event B was 10% of the time between GPS points 2 and 3 (dark grey circles on the time scale), then the distance assigned to walking event B would be 10% of the distance covered between those GPS points (light grey circles of GPS points 2 and 3 on the distance scale). Finally, for rate outcomes, this resulted in the assumption that the rate value applied as a fixed value for the whole time between GPS points. For example, in figure 4, the speed of travel between GPS points 2 and 3 is assumed to be a constant value calculated as the total distance travelled between GPS points 2 and 3 divided by the time between these GPS points.

The type of metric used to summarise an outcome measure (e.g. mean, maximum, sum) to tag the walking event also changed based on the type of GPS outcome. Categorical values were used to tag the walking event with a single categorical value. This was based on converting the duration of the event covered by each type of code, and selecting the one covering the most duration. For example, in figure 2, the parts of walking event C that occurred between the midpoint of GPS points 2 and 3 and the midpoint of GPS points 4 and 5 were categorised as outdoors, and the parts that occur after the midpoint between GPS points 4 and 5 were categorised as indoors. The whole of walking event C was then categorised as outdoors because more of its duration occurred outdoors than indoors. Scale values based on a single GPS point were used to provide an average and a maximum value for each walking event. For example, in figure 3, the average distance from home of walking event A is calculated from the mean of three values, the distance from home of the start of walking event A, the distance from home of GPS point 2, and the distance from home of the end of walking event A. The maximum distance from home of walking event A is the largest of those three points, here the end of walking event A. Scale values based on successive GPS points were used to provide cumulative outcomes. For example, in figure 4, the distance travelled during walking event C is made up of the sum of the distance travelled from the start of walking event C to GPS point 3, the distance travelled between GPS points 3 and 4, and the distance travelled between GPS point 4 and the end of walking event C. Finally, rate values based on successive GPS points were used to provide average values for the walking event. For example, in figure 4, the average speed of walking in walking event C would be the mean of the speed between GPS points 2 and 3, the speed between GPS points 3 and 4, and the speed between GPS points 4 and 5. Note that unlike for distance, as the speed between successive GPS points is assumed to be constant, there is no need to account for how much of walking event C occurs between GPS points 2 and 3 (or between GPS points 4 and 5).

Assessment of Data Integrity Using the Integration Framework

During the assessment period, participants walked an average (mean ± standard deviation) 89±32 minutes per day, in an average of 345 ±102 walking events. GPS data covered 248±95 (72±16%) of walking events, which equated to covering 71±31 min (76±14%) of time spent walking (Table 3). Therefore, however well the event-based physical activity data was integrated with GPS, 24% (by duration) could not be used for integration, due to the absence of GPS data. The remaining information reported here, refers only to those walking events which are covered by GPS data, to explore the integrity of using the framework when GPS data was available.

The number and duration of walking events covered by GPS for each of the types of correspondence with GPS points is shown in Table 3. Walking events corresponding to multiple GPS points, which could be described as the ideal situation, make up only 52% by number but 75% by time of all walking covered by GPS. It is therefore important to retain walking events from other types of correspondence, especially as these might represent particular types of activity. Walking events are typically short with 35% lasting <5 seconds, and the number of walking events taking place entirely between GPS points (38%) was large. However, such events were usually short, representing only 23% of total duration of walking. Walking events taking place entirely between GPS points for which outcome measures could be calculated made up 13% by number and 3% by duration of all walking events. In contrast, walking events taking place entirely between GPS points for which data was missing (i.e. could not be fully calculated, as 0 or 1 external GPS points met inclusion criteria) made up 22% by number and 17% by duration of all walking events. Finally, the case where only a single GPS point was included in the walking event, and no external GPS points met inclusion criteria represented 9% by number and 2% by duration of all walking events covered by GPS data.

Outcome measures relating to a single GPS point could be calculated for more walking events than those relating to the difference between successive GPS points. In total, outcome measures relating to a single GPS point could not be calculated (missing data) for 22% by number and 17% by duration of walking events covered by GPS. Whereas, outcome measures relating to the difference between successive GPS points could not be calculated (missing data) for 35% by number and 22% by duration of walking events covered by GPS.

For walking events taking place entirely between GPS points, relaxing the threshold set for inclusion of external data points gradually increased the number of additional walking events for which location outcomes could be calculated. For example, increasing the allowed duration between GPS points to 120s (from 60s) included an additional 2% of walking events (and an additional 1% of walking duration), and increasing to 180s (from 120s) included a further 2% of walking events and a further 1% of walking duration. However, most of the walking events taking place entirely between GPS points occurred in extended gaps between GPS points. For example, even allowing a gap between GPS points of 30 minutes would still have excluded 12% of events by number (and 10% by walking duration) from analysis. In contrast, for walking events containing only 1 GPS point simply removing the requirement that the walking event covered over 50% of the duration between internal and external GPS points meant that all of these type of walking events would be included in analysis.

Values of rate outcomes between successive GPS points (e.g. speed) were assumed to be a single value for the entire duration between events. Between successive GPS samples, a number of activities could have occurred, for example, a person may have stopped walking, and got in a car and started to drive. This would potentially become more problematic the longer the duration between GPS successive points, for example if the threshold values for inclusion were relaxed. To explore how common such circumstances might be, the distribution of average speed (an example of a rate outcome using successive GPS points) (figure 5), did not differ systematically with different types of correspondence of walking events, and their inclusion or exclusion within analysis using the thresholds for the current study. Therefore, there appeared to be no consistent, structural problem with including currently excluded walking events within the general processing of data to derive GPS based outcomes to add to walking events. However, there were some notable outliers, where average speed was much higher than was feasible for human locomotion, and a threshold (of 3 ms^{-1}) was applied to the average speed to filter out unreasonable cases. In this data and using this threshold, an average of 1±1 walking events per participant of the type entirely between GPS points, and 1±1 walking events per participant of the type with only a single GPS point would have been removed from analysis, which was considered to be acceptable.

During the assessment period, control participants walked for longer than the individuals with intermittent claudication (Table 3). Additionally, a slightly higher proportion of that walking was covered by GPS data in the control participants (82% of the duration of walking) than for participants with intermittent claudication (71% of the duration of walking). However, the distribution of different types of walking events by proportion of walking was similar between the two groups across all categories.

Discussion

A framework was created that allows the consistent integration of event-based and sampled data. In this case, the event-based data represented walking physical activity, and the sampled data represented GPS data. The use of the framework will allow the development of outcomes beyond the capabilities of each method individually, extending to measurement of behaviours derived from activity, speed and trajectories (Jankowska et al., 2015). Although the framework is not limited to physical activity and GPS data, the appropriateness of its use was assessed within the context of physical activity (specifically walking) and GPS derived outcomes.

The framework identified different types of correspondence between walking events and GPS data points. The ideal data set would consist of walking events which included several GPS data points, allowing a well-characterised assessment of the positional context within which that walking event was conducted. However, this type of walking event made up only around half (52%) of all walking events covered by the GPS, although they made up three-quarters (75%) of the duration of walking covered. It was therefore necessary, in order to fully utilise the GPS data, to incorporate other types of walking event into the framework.

Walking events entirely occurring between successive GPS data points constituted most of the rest of the walking events (38%), and nearly a quarter of time spent walking (23%). Including these events within the integration framework was relatively easy, in fact easier for some groupings of GPS outcomes, for example those relating to information between successive GPS data points. However, it has the potential to induce statistical anomalies within the data itself. For example, through lack of other information, rate outcomes between successive data points were applied as a single variable to all walking events occurring between those GPS data points. In extreme cases, this might lead to clustering of identical values which misrepresent the natural variation within the outcome measure. Careful consideration of aggregating the information from individual walking events across days and individuals (for example weighting averages to account for duration of walking rather than number of events), could minimise such effects.

Thresholds were used to determine whether GPS points external to walking events were sufficiently close for their location information to be used for outcome measures for that walking event. The current study used the basic principle that GPS signals that were more than 1-minute apart, or more than 30 seconds from the end of the walking event were inappropriate. This resulted in the exclusion of nearly one third (31%) of all walking events, which represented 19% of all walking, of which 22% were walking events entirely between GPS points and 9% were walking events with only a single GPS point included. All of the walking events with only a single GPS point (i.e. 9%) would have been included in analysis, if the condition that the walking event covered over 50% of the time between the internal and external GPS points. This suggests that these walking events were close in time to other GPS points, but were of much shorter duration than the time between GPS points. This makes sense, as it would be unlikely that there would be a long gap on both sides of a GPS point. On the other hand, increasing the time between GPS points surrounding walking events that corresponded to no internal GPS points did not necessarily result in inclusion of those points. This suggests that these walking events are true missing data, as many of them occurred during long periods when GPS data was not recorded. Such coverage might occur, for example, when walking under dense tree cover or when indoors.

One concern of including data from external GPS points was that, by including rate data from GPS points temporally remote from the walking event, real-world movement, such as change of transport mode to a car journey, may render the rate between GPS points meaningless. Investigating the distribution of speed measured for all types of walking events, suggested that in general the distribution of speed by walking event was similar across the types of walking event and GPS correspondence. Given that all walking events with a single internal GPS point had an external GPS point within 30s of the end of the walking event, it is possible that such walking events could be reasonably included within analysis. However, there were some walking events with abnormally high speeds, which necessitated additional speed threshold to be applied to exclude such unreasonable data. This only removed an average of one event per participant from walking events with a single GPS point so data loss from this additional speed threshold was minimal. The threshold used has not been validated, and it is not clear what value provides the best balance between data use, sensitivity and specificity. It is also possible that other data manipulation techniques could have been used to impute a more appropriate value of speed, as opposed to removing the events from the data.

In the current study, GPS data was available for 72% of walking events and 76% of walking duration measured by the activity monitors, indicating that 24% of activPAL walking was missing GPS data. Other studies have found a smaller percentage of accelerometer data was missing GPS data (17% Meseck et al., 2016; 7% Klinker, Schipperijn, Totfager, Kerr, & Troelsen, 2015) when using an accelerometer which was not worn overnight. It is possible that the longer duration of activity not covered by GPS signal in the current study might be due to the continuous 24-hour (and thus longer) wear time of the activPAL monitor. It is also important to consider that there was likely to be a systematic bias in terms of the times and locations that were not covered by GPS data. For example, if the participant consistently did not switch on the GPS unit unless they were leaving home. This is not an aspect of measurement bias concerning the integration framework, but is a feature of participant compliance and data collection processes. Compliance could potentially be improved by providing robust instructions, reminders and/or incentives to use GPS, and by technological improvements that, for example, allow GPS devices to be switched on continuously for longer periods, removing participant induced issues concerning switching the device on and off (Kerr et al., 2011; Klinker et al., 2015). One option to improve the period for which data can be collected without requiring a change of batteries, would be to sample the GPS data less frequently (Kerr et al., 2011). However, it should be noted that this might affect the relative proportions of the correspondence of GPS data points and walking events, for example increasing the proportion of walking events occurring entirely between GPS data points.

Integration of physical activity data, GPS data and the use of GIS systems has the potential to increase knowledge over above each individually, and allow the answering of new and powerful research questions. Jankowska et al. (2015) published a framework for combining GPS, GIS, and accelerometer data, exploring the strengths and limitations of the data sources, and identifying the enhanced research questions which could be answered by using combined data. It was acknowledged that appropriate methods were required to process and combine data from different sources. However, Jankowska et al. (2015) assumed that physical activity data was collected as epochs and counts, and therefore the current framework complements and extends that work, allowing questions suitable for investigation by event-based physical activity to be evaluated.The strength of this framework is that it is comprehensive, and could be used in a number of different situations to integrate any event-based and sampled data. However, care would need to be taken to consider, and check, that the information used is appropriate. For example, different distributions of types of events or sampling rate may change the relative contribution of types of event to data outcomes, which may in turn skew variables derived from those outcomes. Additionally, the framework could be expanded for other types of outcomes, e.g. rate data based on a single point (with reference to a fixed/point, rate).

The framework was tested on data collected from older adults (some of whom had a chronic condition affecting their mobility), and thus may not be generalisable to a younger or more active population. Participants with four days of GPS data were included in this analysis, however longer data collection periods (e.g. 14 days) are required to adequately capture some types of activity space use (Holliday, Howard, Emch, Rodríguez, Rosamond, & Evenson, 2017; Zenk, Matthews, Kraft, & Jones, 2018). The integration framework makes a number of assumptions, which appear appropriate for this data, but which may not be generally appropriate. The framework treats all outcomes as occurring linearly with duration between GPS points. This is almost certainly not true, but with GPS sampled every 5 seconds apart, the effect on data processing is negligible. However, the assumption might not hold for lower GPS sampling rates (e.g. to increase battery life) or for data evaluating fast changing behaviours. In the current data, a change of mode of travel, for example getting in a car, could potentially heavily influence some GPS derived outcomes (e.g. speed). In the current study we found this only affected an average of one of these events per person, but we did have to introduce an extra threshold to remove those occasional cases, to prevent them from skewing the data. It was assumed that the activity monitor and GPS device were completely synchronised across the whole week. In practice this may not be the case, for example a temporal drift of 0.6 seconds per day has been shown between activPAL and GPS devices (Steel, Bejarno, & Carlson, 2019). Synchronous analysis of GPS and physical activity data assumes that both devices were worn or carried at all times. Although diaries were used to remove data where the participant reported removing the activPAL monitor or forgetting to carry the GPS device, no data driven method was used to verify this. Manufacturer information and preliminary testing indicated a positional accuracy for the GPS device of 10m. However, no formal validation or reliability testing has been conducted for the GPS device, and thus it is not clear how suitable this device was for use compared to other available GPS devices. In the current study, the GPS data was cleaned using the methods described here. Cleaning of GPS data is necessary, and some large studies provide standardised methods, however for this monitor, no such standards existed. This is a separate consideration from the use of the integration framework, but may have affected data quality.

Conclusions

Integration of objectively measured physical activity with contextual measures (e.g. GPS) has the potential to enhance interpretation and use of both types of evidence. Event-based physical activity data represents the potential to assess patterns of movement. A framework for the integration of event-based (physical activity) and sampled (GPS) data was developed, providing different methods of calculation of GPS derived outcome measures for each walking event based on the type of GPS outcome and the correspondence of that data with the walking event. This framework could also potentially be used to integrate data of similar types in other situations.

Figure Captions

Figure 1. Schematic diagram showing the correspondence between walking events and GPS points, and rules for including or excluding GPS points from walking events. The dark grey circles 1 to 5 represent the GPS data points on the time scale. Rectangles A to C represent the duration of walking events on the time scale. Three categories of walking events are shown: containing zero GPS points/between GPS points (walking event B, between GPS data points 2 and 3); containing at least one GPS point (walking event C, containing GPS data points 3 and 4); and containing exactly one GPS point (walking event A, containing GPS data point 2). The rules for inclusion of GPS points with corresponding walking events are shown schematically at the top of the figure. For walking events containing zero GPS points, the duration between external GPS points 2 and 3 was less than 60s, and therefore both were included for calculation of outcome measures. For walking events containing one or more GPS points, GPS point 1 met the criteria for walking event A extending over 50% of the time between GPS points 1 and 2, but was more than 30 seconds from the end of walking event A,

and was thus not included. GPS point 5 met both criteria and was included as part of walking event C.

Figure 2. Schematic diagram showing the tagging of walking events with categorical data derived from a single GPS point (in this example, whether the GPS point was recorded while indoors or outdoors). The dark grey circles 1 to 5 represent the GPS data points on the time scale. Rectangles A to C represent the duration of walking events on the time scale. Changes in category are assumed to occur at the mid-point (50%) in time between GPS points, and a single categorical value is derived for each walking event based on the category with the largest duration during that event.

Figure 3. Schematic diagram showing the tagging of walking events with scale data derived from a single GPS point (in this example, distance from home). The dark grey circles 1 to 5 represent the GPS data points on the time scale and the light grey circles 1-5 represent those same GPS data points on the distance from home scale. Rectangles A to C represent the duration of walking events on the time scale. The rate of change of distance from home between GPS points is assumed to be linear, and is scaled with respect to the relative duration of walking event and GPS point. Maximum and mean values of the outcome measure are derived for each walking event.

Figure 4. Schematic diagram showing the tagging of walking events with scale data derived from two successive GPS points (in this case, distance moved). The dark grey circles 1 to 5 represent the GPS data points on the time scale and the light grey circles 1-5 represent those same GPS data points on the distance scale. Rectangles A to C represent the duration of walking events on the time scale. The rate of change of distance between GPS points is assumed to be linear, and is scaled with respect to the relative duration of walking event and GPS point. A cumulative value for the outcome measure is derived from each walking event.

Figure 5. Distribution of the speed of each walking event by type of walking event (zero internal GPS points; one or more internal GPS points) and whether they were included or excluded, as a percentage of total event of each type.

References

Bohannon R.W., Andrews A.W. (2011) Normal walking speed: a descriptive meta-analysis. *Physiotherapy,* 97, 182-189.

Carlson J.A., Schipperijn J., Kerr J., Saelens B.E., Natarajan L., Frank L.D., … Sallis J.F. (2016) Locations of physical activity as addressed by GPS in young adolescents. *Pediatrics,* 137(1), 1-10.

Chastin S.F.M., Dall P.M., Tigbe W.W., Grant P.M., Ryan C.G., Rafferty D., Granat M.H. (2009) Compliance with physical activity guidelines in a group of UK based postal workers using an objective monitoring technique. *European Journal of Applied Physiology,* 106(6), 893-899.

Chastin S.F.M, Dontje M.L., Skelton D.A., Čukić I., Shaw R.J., Gill J.M.R., … Dall P.M. on behalf of the Seniors USP. (2018) Team Systematic comparative validation of self-report measures of sedentary time against an objective measure of postural sitting (activPAL). *International Journal of Behavioral Nutrition and Physical Activity,* 15, 21.

Clarke C.L., Holdsworth R.J., Ryan C.G., & Granat M.H. Free-living physical activity as a novel outcome measure in patients with Intermittent Claudication. (2013) *European Journal of Vascular and Endovascular Surgery,*45(2), 162-167.

Dall P.M., McCrorie P.R.W., Granat M.H., & Stansfield B.W. (2013) Step accumulation per minute epoch is not the same as cadence for free-living adults. *Medicine & Science in Sports & Exercise,* 45(10), 1995-2001.

Evenson K.R., Wen F., Hillier A., & Cohen D.A. (2013) Assessing the contribution of parks to physical activity using GPS and accelerometry. *Medicine & Science in Sports & Exercise,* 45(10), 1981-1987.

Granat M.H. (2012) Event-based analysis of free-living behaviour. *Physiolical Measurement,* 33, 1785-1800.

Garg P.K., Tian L., Criqui M.H., Liu K., Ferrucci L., Guralnik J.M., Tan J., & McDermott M.M. (2006) Physical activity during daily life and mortality in patients with peripheral arterial disease. *Circulation,* 114, 242-248.

Grant P.M., Dall P.M., Mitchell S.L., & Granat M.H. (2008) Activity monitor accuracy in measuring step number and cadence in community-dwelling older adults. *Journal of Aging and Physical Activity,* 16(2), 201-214.

Grant P.M., Ryan C.G., Tigbe W.W., & Granat M.H. (2006) The validation of a novel activity monitor in the measurement of posture and motion during everyday activities. *British Journal of Sports Medicine*, 40(12), 992-997.

Goel R., Gani S., Guttikunda S.K., Wilson D., Tiwari. (2015) On-road $PM_{2.5}$ pollution exposure in multiple transport microenvironments in Delhi. *Atmospheric Environment*, 123, 129-138.

Holliday K.M., Howard A.G., Emch M., Rodríguez D.A., Rosamond W.D., & Evenson K.R. (2017) Deriving a GPS monitoring time recommendation for physical activity studies of adults. *Medicine and Science in Sports and Exercise*, 49(5), 939-947.

Hurvitz P.M., Moudon A.V., Kang B., Saelens B.E., & Duncan G.E. (2014) Emerging technologies for assessing physical activity behaviours in space and time. *Frontiers in Public Health,* 2, 2.

James P., Jankowska M., Marx C., Hart J.E., Berrigan D., Kerr J., … Laden F. (2016) “Spatial Energetics” integrating data from GPS, accelerometry, and GIS to address obersity and inactivity. *American Journal of Preventive Medicine,* 51(5), 792-800.

Jankowska M.M., Schipperijn J., & Kerr J. (2015) A framework for using GPS data in physical activity and sedentary behaviour studies. *Exercise Sports Science Review,* 43(1), 48-56.

Kerr J., Duncan S., & Schipperijn J. (2011) Using global Positionng Systems in health research. A practical approach to data collection and processing. *American Journal of Preventive Medicine,* 41(5), 532-540.

Kim Y., Beets M.W., Pate R.R., & Blair S.N. (2013) The effect of reintegrating Actigraph accelerometer counts in preschool children: Comparison using different epoch lengths. *Journal of Science and Medicine in Sport,* 16(2), 129-134.

Klinker C.D., Schipperijn J., Totfager M., Kerr J., & Troelsen J. (2015) When cities move children: Development of a new methodology to assess context-specific physical activity behaviour among children and adolescents using accelerometers and GPS. *Health & Place,*, 90-99.

Kyu H.H., Bachman V.F., Alexander L.T., Mumford J.E., Afshin A., Estep K., … Forouzanfar M.H. (2016) Physical activity and risk of breast cancer, colon cancer, diabetes, ischemic heart disease, and ischemic stroke events: systematic review and dose-response meta-analysis for the Global Burden of Disease Study 2013. *BMJ,* 354, i3857.

Lee I.M., Shiroma E.J., Lobelo F., Puska P., Blair S.N., & Katzmarzyk P.T. (2012) Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *Lancet*, 380, 219-229.

Lythgo N., Wilson C., Galea M. (2011) Basic gait and symmetry measures for primary school-aged children and young adults. II: walking at slow free and fast speed. *Gait & Posture,* 33, 29-35.

Matthews C.E. (2005) Calibration of accelerometer output for adults. *Medicine & Science in Sports & Exercise*, 37, 5512-5522.

Meseck K., Jankowska M.M., Schipperijn J., Natarajan L., Godbole S., Carlson J., … Kerr J. (2016) Is missing geographic positioning system data in acceleromterty studies a problem, and is imputation the solution? *Geospatial Health,* 11(2), 403.

Norgren L., Hiatt W.R., Dormandy J.A., Nehler M.R., Harris K.A., & Fowkes F.G. (2007) Inter-society consensus for the management of peripheral arterial disease (TASC II). *Journal of Vascular Surgery,* 45(1), S5-67.

Rafferty D., Dolan C., & Granat M. (2016) Attending a workplace: its contribution to volume and intensity of physical activity. *Physiological Measurement*, 37(12), 2144-2153.

Saunders D.H., Sanderson M., Hayes S., Kilrane M., Greig C.A., Brazzelli M., & Mead G.E. (2016) Physical fitness for stroke patients. *Cohrane Database of Systematic Reviews,* 2016(3), CD003316.

Strain T., Milton K., Dall P.M., Standage M., & Mutrie N. (2019) How are we measuring physical activity and sedentary behaviour in the four home nations of the United Kingdom? A narrative review of current surveillance measures and future directions. *British Journal of Sports Medicine* (e-pub ahead of print).

Steel C., Bejarno C., & Carlson J.A. (2019) Time drift considerations when using GPS and accelerometers. *Journal of the Measurement of Physical Behaviour,* 2, 203-207.

Wing M.G., Eklund A., & Kellogg L.D. (2005). Consumer-grade global positioning system (GPS) accuracy and reliability. *Journal of Forestry*, 103, 169-173

Zenk S.N., Matthews S.A., Kraft A.N., & Jones K.K. (2018). How many days of global positioning system (GPS) monitoring do you need to measure activity space environments in health research?. *Health & Place*, 51, 52-60.

Table 1: Types of outcome measure that can be derived from GPS data, with examples derived for the current study.

GPS points used to derive outcome	Type of variable	outcome name	description of calculation of outcome for the current study
derived from a single GPS point	Categorical	Home/Away location code	The GPS point was coded as home if distance from home was less than a threshold (in the current study we used <=50m), and away from home if above the threshold.
		Indoor/Outdoor location code	The GPS point was coded as indoors if the signal to noise ratio was lower than a threshold value (in the current study we used <=212 dB), and outdoors if below the threshold.
	Scale	Distance from home	calculated using Pythagoras' theorem based on the difference in eastings and northings (in metres) of the GPS point from a fixed reference designated as the home location. In the current study the easting and northing values for the home location of the participant, were derived from identifying the central location of the house outline on a MasterMap in ArcMap (v9.3).
derived from successive GPS points	Categorical	Gradient Code	The GPS point was coded as uphill, flat or downhill if the gradient value with respect to the previous GPS point was above, below or between threshold values (in the current study we used ±1.3%).
	Scale	Distance travelled	calculated using Pythagoras' theorem as the square root of the square of the differences in easting and northing values (unit of metres) between two successive GPS data points.
		Height change	the difference in value between the values of height above sea level of successive GPS points. A positive value was considered height climbed and a negative vale was considered height lost.
	Rate	Speed	calculated as the distance travelled between successive GPS points divided by the time difference between those two points.
		Gradient	calculated as the height difference between successive GPS points, divided by the distance between those points.

Table 2: Framework for integrating GPS data with event-based physical activity data.

<table>
<tr><th colspan="2" rowspan="2">Type of walking event
(in terms of correspondence GPS points)</th><th colspan="5">Type of GPS Outcome Measure</th></tr>
<tr><th colspan="2">Outcomes relating to a single GPS point</th><th colspan="3">Outcomes relating to difference between successive GPS points</th></tr>
<tr><td># internal GPS points (within duration of walking event)</td><td># external GPS points meeting inclusion criteria</td><td>Categorical

e.g. location category: Home/Away; Indoor/Outdoor</td><td>Scale (from fixed reference)

e.g. distance from Home</td><td>Categorical

e.g. Gradient Category: Uphill/Downhill/Flat</td><td>Scale

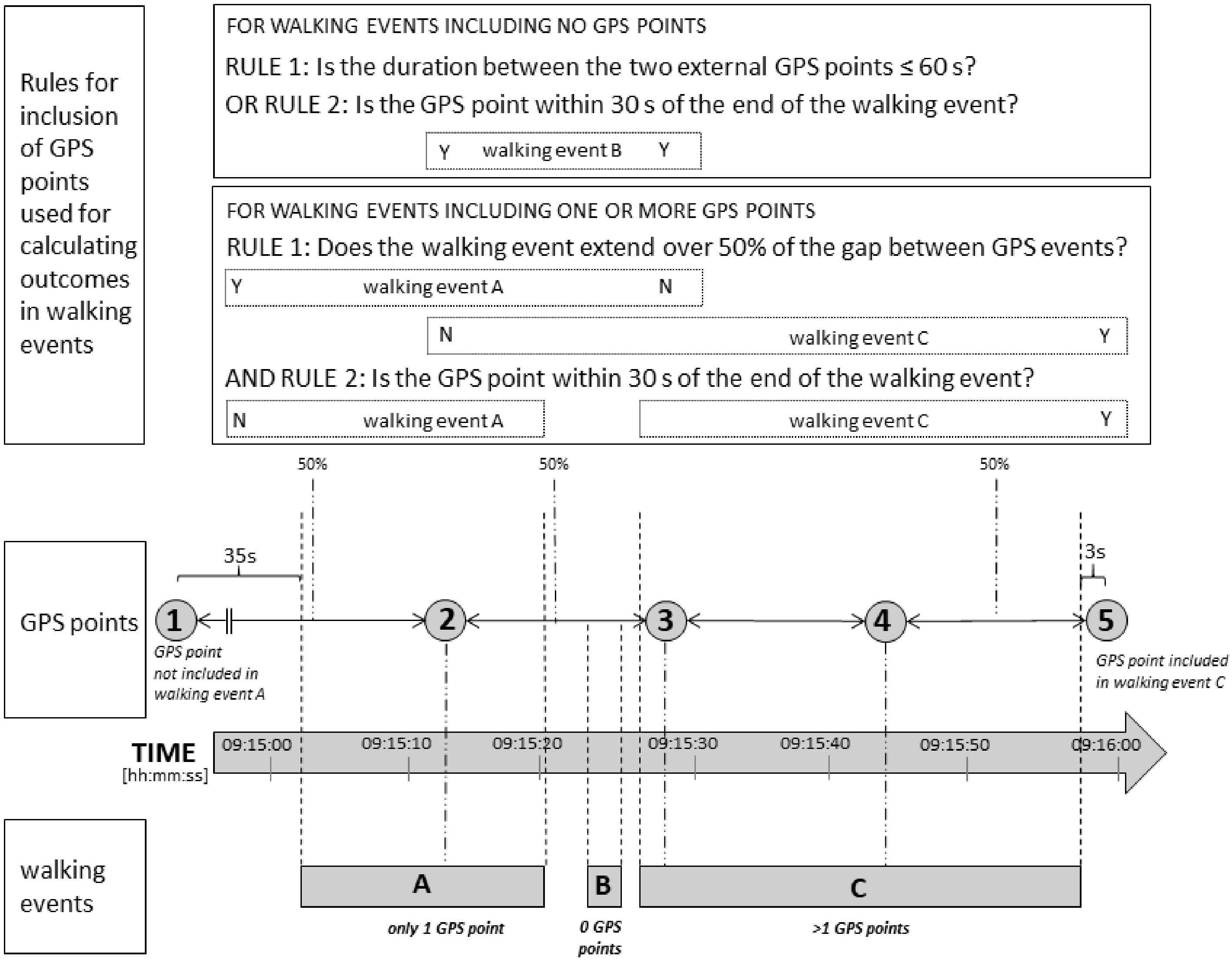
e.g. Distance; Height Climbed/ Descended</td><td>Rate

e.g. Speed; Gradient (value)</td></tr>
<tr><td rowspan="3">0</td><td>0</td><td>missing</td><td>missing</td><td>missing</td><td>missing</td><td>missing</td></tr>
<tr><td>1</td><td rowspan="2">Categorical code for the walking event is taken from one GPS point either before or after, depending on which one > 50% of the walking event is closer to in time</td><td rowspan="2">Assume the scale value changes at a steady rate between the GPS points on either side. Mean outcome value is the outcome value at the mid-point of the walking event (scaled by duration). The maximum outcome value is the largest value of the start and end of the walking event (scaled by duration).</td><td>missing</td><td>missing</td><td>missing</td></tr>
<tr><td>2</td><td>Categorical code for the walking event is based on gradient between GPS points on each side of the walking event.</td><td>Outcome value for the walking event is a proportion (scaled by duration) of the value between GPS points each side of the walking event.</td><td>Outcome value for the walking event is that calculated from GPS points each side of the walking event</td></tr>
<tr><td rowspan="2">1</td><td>0</td><td rowspan="3">Each GPS point is coded. The code is assumed to extend half way (in time) to the previous/next GPS point. Duration of walking event covered by each code is calculated. Walking event is coded as categorical code with greatest proportion of event covered.</td><td rowspan="3">Mean outcome value for the walking event is the sum of the outcomes from all points within the walking event, plus proportional values (based on duration) of the start/end of the walking event extending past first/last internal GPS point, if included, divided by number of points. The maximum outcome value is the largest value of the included GPS points or the value at the start/end of the walking event.</td><td>missing</td><td>missing</td><td>missing</td></tr>
<tr><td>1 or 2</td><td rowspan="2">The categorical code applies to time between pairs of GPS points. The duration of walking event covered by each code is calculated. Walking event is coded as categorical code with greatest proportion of event covered.</td><td rowspan="2">Outcome value for the walking event is the sum of outcome values between all GPS points inside the walking event, plus proportional values (based on duration) of part of walking event that extends past the first and last internal GPS points</td><td rowspan="2">Outcome value for the walking event is the sum of all outcome values between successive GPS points in the event (including outcome for points before and after event if they meet the criteria for inclusion) divided by the number of points</td></tr>
<tr><td>>1</td><td>0, 1 or 2</td></tr>
</table>

Table3: Number and duration of walking events by correspondence with GPS points

<table>
<tr><th colspan="2"></th><th colspan="2">Whole Group</th><th colspan="2">Individuals with Intermittent Claudication</th><th colspan="2">Control older adults</th></tr>
<tr><th colspan="2"></th><th>Number of Walking Events [n/day] (%)</th><th>Total Time Spent Walking [min/day] (%)</th><th>Number of Walking Events [n/day] (%)</th><th>Total Time Spent Walking [min/day] (%)</th><th>Number of Walking Events [n/day] (%)</th><th>Total Time Spent Walking [min/day] (%)</th></tr>
<tr><td colspan="2">All Walking Events[1]</td><td>345 ± 102</td><td>89 ± 32</td><td>315 ± 112</td><td>72 ± 28</td><td>374 ± 83</td><td>106 ± 28</td></tr>
<tr><td colspan="2">All Walking Events covered by GPS[2]</td><td>248 ± 95
(72 ± 16)</td><td>71 ± 31
(76 ± 14)</td><td>207 ± 102
(65 ± 18)</td><td>53 ± 22
(71 ± 15)</td><td>288 ± 68
(77 ± 11)</td><td>88 ± 28
(82 ± 10)</td></tr>
<tr><td colspan="2">Type of walking event</td><td colspan="6" rowspan="2"></td></tr>
<tr><td># internal GPS points (within duration of walking event)</td><td># external GPS points meeting inclusion criteria</td></tr>
<tr><td rowspan="3">0</td><td>0[3]</td><td>51 ± 36
(22 ± 15)</td><td>11 ± 8
(17 ± 11)</td><td>47 ± 38
(24 ± 15)</td><td>10 ± 7
(19 ± 11)</td><td>54 ± 33
(20 ± 14)</td><td>12 ± 8
(14 ± 11)</td></tr>
<tr><td>1[3]</td><td>9 ± 6
(4 ± 2)</td><td>2 ± 2
(3 ± 2)</td><td>7 ± 3
(4 ± 2)</td><td>2 ± 1
(3 ± 1)</td><td>12 ± 7
(4 ± 3)</td><td>2 ± 1
(3 ± 2)</td></tr>
<tr><td>2[3]</td><td>32 ± 16
(13 ± 3)</td><td>2 ± 1
(3 ± 1)</td><td>26 ± 16
(12 ± 4)</td><td>1 ± 1
(3 ± 2)</td><td>37 ± 14
(13 ± 3)</td><td>2 ± 1
(3 ± 1)</td></tr>
<tr><td rowspan="2">1</td><td>0[3]</td><td>23 ± 12
(9 ± 3)</td><td>1 ± 1
(2 ± 1)</td><td>19 ± 13
(9 ± 3)</td><td>1 ± 1
(2 ± 1)</td><td>27 ± 11
(9 ± 2)</td><td>2 ± 3
(2 ± 1)</td></tr>
<tr><td>1 or 2[3]</td><td rowspan="2">133 ± 69
(52 ± 12)</td><td rowspan="2">55 ± 27
(75 ± 13)</td><td rowspan="2">108 ± 70
(51 ± 12)</td><td rowspan="2">39 ± 19
(72 ± 12)</td><td rowspan="2">158 ± 59
(54 ± 12)</td><td rowspan="2">70 ± 28
(77 ± 12)</td></tr>
<tr><td>>1</td><td>0, 1 or 2[3]</td></tr>
</table>

Data presented as groups mean ±standard deviation (% of total) for all participants. [1]represents the denominator for the percentage, so percentage not presented. [2]presented as percentage of all walking events. [3]presented as percentage of walking events covered by GPS.



Rules for inclusion of GPS points used for calculating outcomes in walking events
FOR WALKING EVENTS INCLUDING NO GPS POINTS
RULE 1: Is the duration between the two external GPS points ≤ 60 s?
OR RULE 2: Is the GPS point within 30 s of the end of the walking event?
Y walking event B Y
FOR WALKING EVENTS INCLUDING ONE OR MORE GPS POINTS
RULE 1: Does the walking event extend over 50% of the gap between GPS events?
Y walking event A N
N walking event C Y
AND RULE 2: Is the GPS point within 30 s of the end of the walking event?
N walking event A
walking event C Y
50%
50%
50%
35s
3s
GPS points
1
2
3
4
5
GPS point not included in walking event A
GPS point included in walking event C
TIME
[hh:mm:ss]
09:15:00
09:15:10
09:15:20
09:15:30
09:15:40
09:15:50
09:16:00
walking events
A
B
C
only 1 GPS point
0 GPS points
>1 GPS points

Figure 1

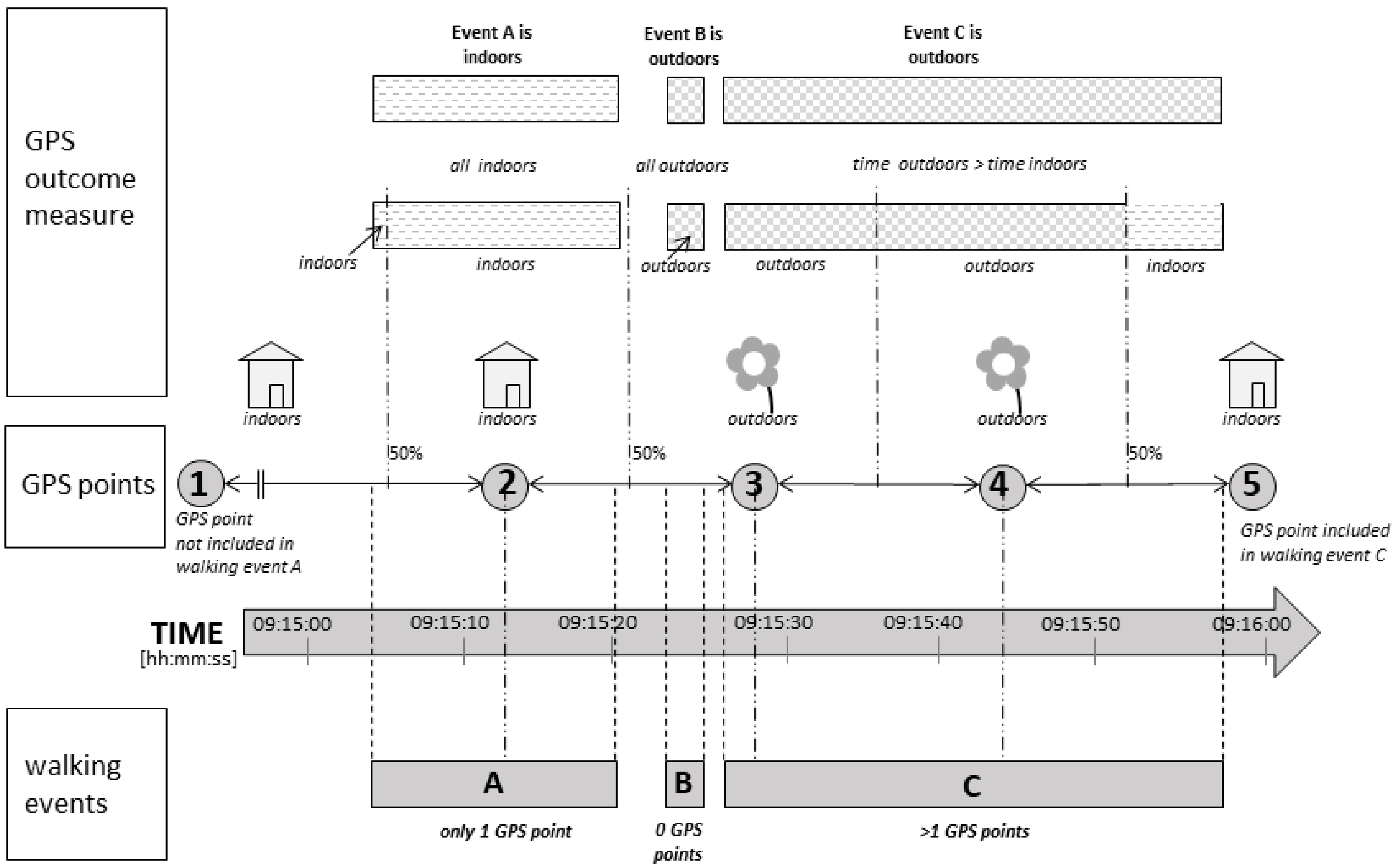

GPS outcome measure
Event A is indoors
Event B is outdoors
Event C is outdoors
all indoors
all outdoors
time outdoors > time indoors
indoors
indoors
outdoors
outdoors
outdoors
indoors
indoors
indoors
outdoors
outdoors
indoors
50%
50%
50%
GPS points
1
2
3
4
5
GPS point not included in walking event A
GPS point included in walking event C
TIME [hh:mm:ss]
09:15:00
09:15:10
09:15:20
09:15:30
09:15:40
09:15:50
09:16:00
walking events
A
B
C
only 1 GPS point
0 GPS points
>1 GPS points

Figure 2

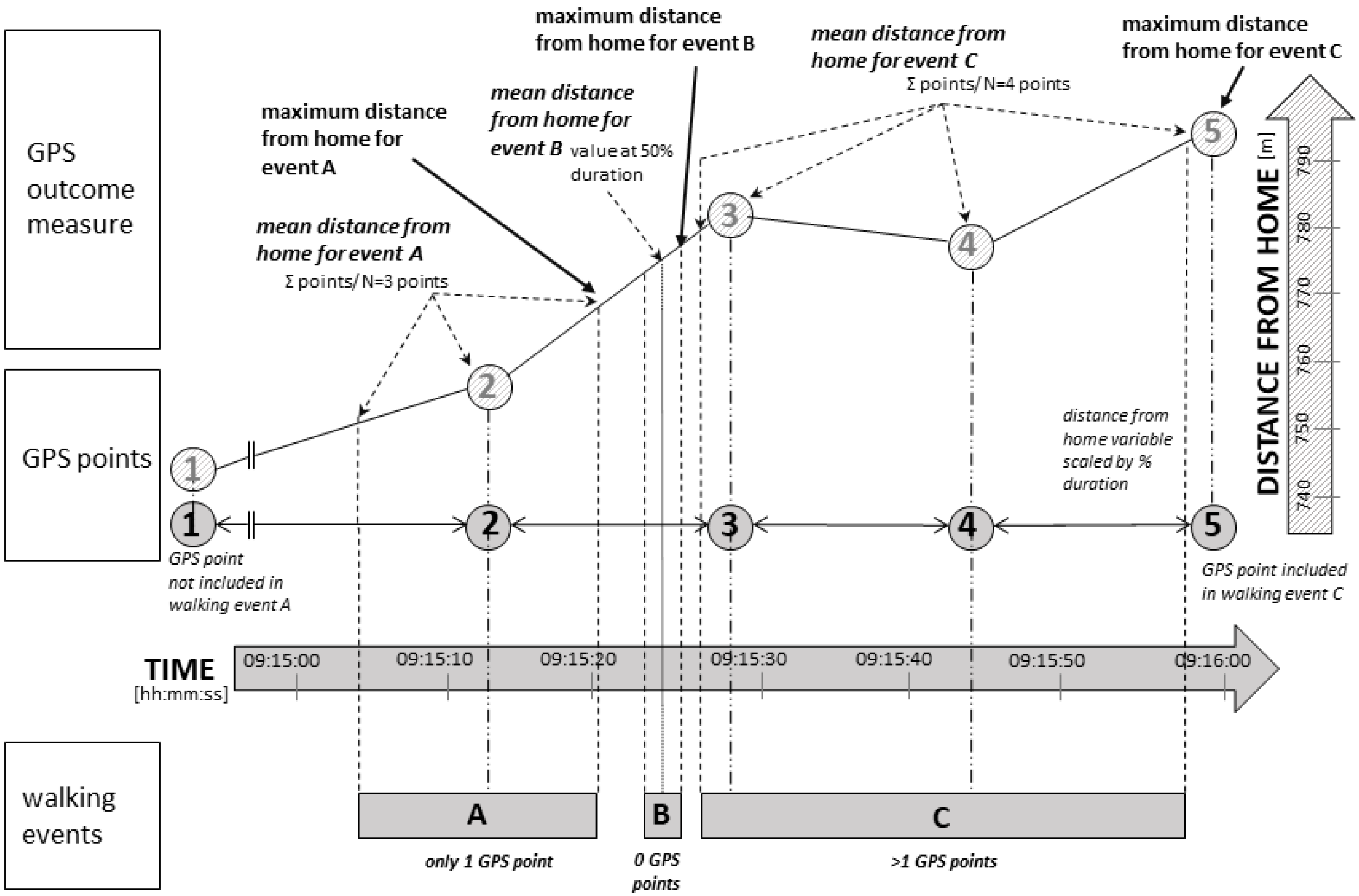
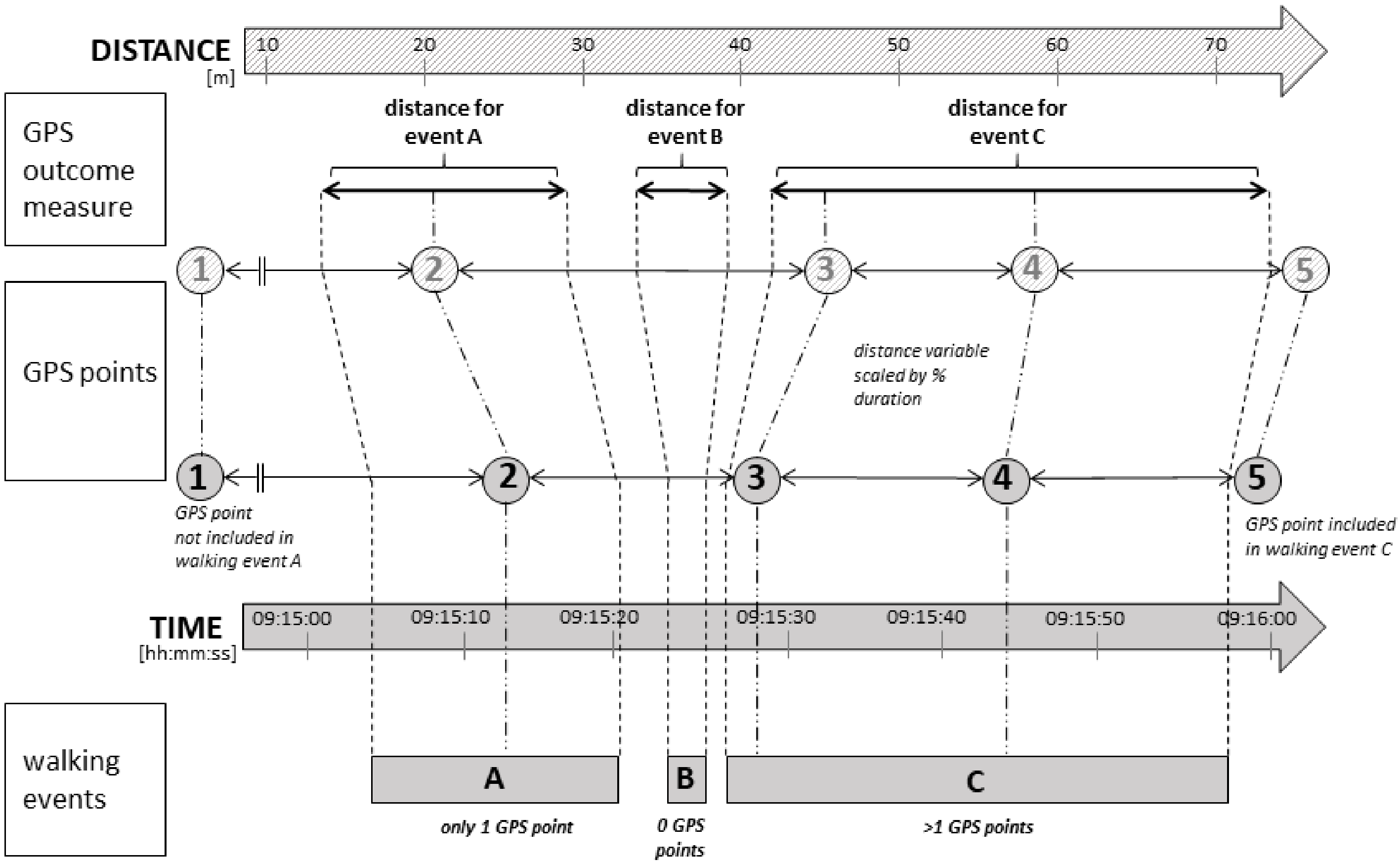

GPS outcome measure
GPS points
walking events
maximum distance from home for event A
mean distance from home for event A
Σ points/ N=3 points
mean distance from home for event B
value at 50% duration
maximum distance from home for event B
mean distance from home for event C
Σ points/ N=4 points
maximum distance from home for event C
DISTANCE FROM HOME [m]
740
750
760
770
780
790
distance from home variable scaled by % duration
1
2
3
4
5
GPS point not included in walking event A
GPS point included in walking event C
TIME [hh:mm:ss]
09:15:00
09:15:10
09:15:20
09:15:30
09:15:40
09:15:50
09:16:00
A
B
C
only 1 GPS point
0 GPS points
>1 GPS points

Figure 3



DISTANCE
[m]
10
20
30
40
50
60
70
GPS outcome measure
distance for event A
distance for event B
distance for event C
GPS points
1
2
3
4
5
distance variable scaled by % duration
GPS point not included in walking event A
GPS point included in walking event C
TIME
[hh:mm:ss]
09:15:00
09:15:10
09:15:20
09:15:30
09:15:40
09:15:50
09:16:00
walking events
A
B
C
only 1 GPS point
0 GPS points
>1 GPS points

Figure 4

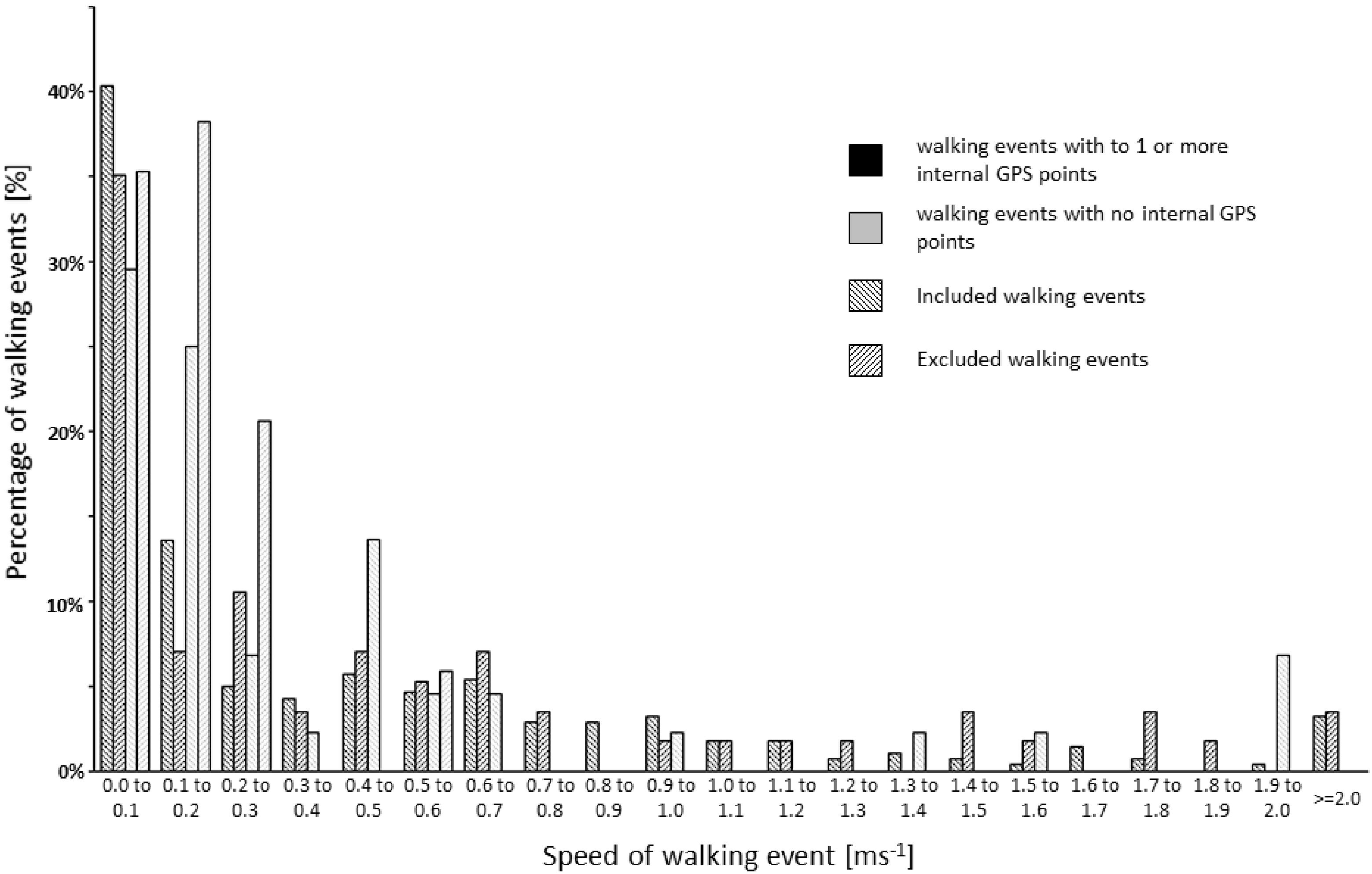

40%
30%
20%
10%
0%
Percentage of walking events [%]
walking events with to 1 or more internal GPS points
walking events with no internal GPS points
Included walking events
Excluded walking events
0.0 to 0.1
0.1 to 0.2
0.2 to 0.3
0.3 to 0.4
0.4 to 0.5
0.5 to 0.6
0.6 to 0.7
0.7 to 0.8
0.8 to 0.9
0.9 to 1.0
1.0 to 1.1
1.1 to 1.2
1.2 to 1.3
1.3 to 1.4
1.4 to 1.5
1.5 to 1.6
1.6 to 1.7
1.7 to 1.8
1.8 to 1.9
1.9 to 2.0
>=2.0
Speed of walking event [ms-1]

Figure 5