FINAL REPORT

24/01/25

GARMIN VIVOACTIVE 5 - GPS AND HEALTH DATA QUALITY

DR PAUL MCCRORIE
DR FIONA CARYL

MRC/CSO Social and Public Health Sciences Unit

EXECUTIVE SUMMARY

Wearable technology can be a valuable addition to large scale, cohort based, surveys where participant burden is low, and data can be collected passively. Commercially available Smartwatches and 'fitness trackers' have become increasingly popular and combine data from multiple internal sensors to present users with a holistic representation of their health and wellbeing.

The Garmin Vivoactive 5 has been chosen by the Understanding Society team at the University of Essex with the long-term plan to deploy within their Innovation Panel. The Understanding Society team has employed Avicenna Research - a Canadian based tech company that specialises in converting smartphone/wearable data for applied research - to develop a system (app development, data collection, storage, extraction, and presentation) that captures health behaviour (physical activity, PA), sleep, and spatial data. Prior to integration into a larger, and costly, data collection the device should be rigorously tested to ensure it meets the requirements of the team and the Understanding Society team have consulted the University of Glasgow to assist with testing the solution produced from Avicenna.

The key deliverables of this consultancy were:

1. A test of the Garmin Vivoactive 5 capability to operate independently of, and in conjunction with, users' smartphones to collect locational data (i.e., GPS X/Y coordinates)
2. A test of **static** accuracy of the Garmin Vivoactive 5's GPS collection (how well does it measure users' location if residing in a known, non-changing, location).
3. A test of **dynamic** accuracy of the Garmin Vivoactive 5's GPS collection (how well does the device measure the locational data of someone when in a free-living context).
4. Headline descriptives of the physical activity, sedentary, and sleep variables extracted by Avicenna.
5. Limitations, recommendations, and potential usability of the underlying data for scientific research

Findings

Physical activity

- Summary level data presented as daily level of physical activity have some research value, however the available variables need to be expanded to gain optimal use.

- As a priority, Raw accelerometry data should be requested in future iterations of the solution provided by Avicenna to ensure The University of Essex/Understanding Society team have full access and control over the underlying data, free from proprietary processing.
- Where possible, Garmin processed summary data at 1-minute epochs should be pursued, allowing for linkage to GPS data.

GPS

- Protocol trials evaluated GPS data quality across three device configurations: smartwatch-only (P1: Smartphone Present, Smartwatch Absent), smartphone-only (P2: Smartphone Present, Smartwatch Absent), and dual-device (P3: Both Devices Present), assessing GPS data completeness, positional uncertainty, positional accuracy, and spatial adherence during static and moving periods.
- Dual-device configuration (P3) demonstrated optimal performance for health-location studies, achieving best route adherence (6-8m) and consistent static positioning (8-9m). While not maximising data completeness (50-76%), it balances spatial accuracy with essential health metric collection.
- Watch-only tracking (P1) showed adequate movement tracking but reduced precision static positioning (361-466m deviation). Phone-only configuration (P2) achieved highest completeness (62-94%) but lacks ability to record required health metrics.
- Key limitations include poor indoor positioning across all configurations and reduced quality during journey initialisation. Individual and environmental factors significantly influence performance but were beyond the scope of this trial to quantify.
- Recommendations prioritise dual-device deployment with enhanced GNSS quality metrics (DOP, SNR) for improved data quality assessment and validation. Implementation requires GPS initialisation periods, geofencing for indoor detection, and optimisation of epoch aggregation methods, such as those used for this report, which can be enhanced with additional DOP and SNR parameters.
- Future development should focus on acquiring additional GNSS parameters to strengthen position validation and indoor/outdoor transition detection.

General reflections from protocol and wearing the device

Having used several devices in research applications, including large scale data collections it is worthwhile presenting some initial thoughts on the experience. These are summarised in the bulleted points below:

Positives

- Device is waterproof and has been worn swimming and bathing on several occasions. Extremely valuable for compliance levels and minimising burden on participants.
- Exceptionally easy to wear, lightweight, modern, and easy to charge. This will improve compliance in any research protocol.

Negatives

- Onboarding instructions from Avicenna were poor and confusing. This resulted in a few users not being able to contribute to the protocol testing or having zero data collected. Clear communication and guidance are essential if rolled out to the Innovation Panel sample or there is a large risk of non-compliance, poor or even zero data.

Uncertainties to explore

- Variation exists in the battery life cycle between charges, and this should be explored to optimise before roll out. Variation is likely a reflection of two things:
    1. **Type of individual**: The 'standard/average' user is unlikely to be using the device to record multiple walks/runs/other activities and will use the device to record passively in the background. Their battery usage will be minimal and can expect to have up to 7/8 days of battery life before charging. However, you will have the 'right tail' of the population who may use the device more intensively, recording walks/runs/cycling etc. This will drain the battery faster and estimates of time between charges will vary.
    2. **Type/age of smartphone and Bluetooth connectivity**: The Avicenna solution is designed whereby the smartphone will take on the responsibility of capturing GPS data but will defer to the smartwatch (user needs to activate this) when the watch is out of the Bluetooth range of the device. Feedback has suggested that users have had to collect considerable amount of GPS data on their watch as a result of poor Bluetooth connectivity.
- Variation exists in the way smartphone operating systems allow background recording of data. Subtle differences between Android and IOS for instance mean that Apple users must keep the Avicenna app open in the background and cannot swipe this up (close it) or Avicenna loses the ability to collect GPS data from the device. This may have contributed to variation in GPS data collected (see GPS section).

- The smartwatch is designed to provide real-time feedback to users across multiple outcomes of interest. It can also be used to purposefully track specific activities (walks/runs/swimming etc). The functionality, as a commercial device, is of high quality and incredibly useful for most users. However, for research studies, where the primary focus is not behaviour change, there is a high risk of introducing a variety of behaviour change techniques (e.g., real-time feedback, prompts/cues, goal setting, self-monitoring, health info) into the study, thereby potentially artificially inflating the outcomes, and biasing the findings. The implications should be considered carefully and, if possible, action should be taken to minimise the watch functionality. In principle, the device should act passively in the background with little user engagement.

HEALTH DATA – PHYSICAL ACTIVITY

The Garmin Vivoactive 5 is equipped with several sensors that can be utilised to extract information on a user's levels of physical activity, including an accelerometer and heart rate monitor. A common representation of physical activity is through the expression of time spent in certain intensities, and reflects the underlying effort required to perform an activity. Many, if not all, global recommendations for health enhancing PA are largely based on aerobic activity (but do incorporate muscle-strengthening and functional fitness) and often use Moderate-Vigorous intensity physical activity – activity requiring the effort of between 3 and >6 times than that when the body is at rest (Metabolic equivalents) – as the threshold through which greatest health enhancing activity occurs.

The Vivoactive 5 smartwatch calculates the time you spend in MVPA by analysing your heart rate data. When your heart rate exceeds a threshold (70-80% and 80-90% of maximum heart rate for moderate and vigorous respectively) the device records these periods independently. If heart rate monitoring is disabled, the watch estimates moderate-intensity minutes by evaluating your steps per minute (not disclosed but wider literature suggests around 100 steps per min).

Two main datasets are currently available as standard through the Avicenna Dashboard:

i) Daily level, accumulated data presented for several outcome variables; and
ii) 15-minute epoch level data, sub categorised by activity type across each 15-minute epoch.

DAILY DATA

The daily data corresponds to the Garmin Connect's "My Day" section and has been extracted as a snapshot/headline summary of user's health. As these data have been extracted as standard, there is likely some variables that are of little value. The research team will need to query and prioritise outcomes that they think will have both immediate (research team) and wider value, as well as forecast what may be of interest in some future time. As these are standard outcomes from the device, there is an argument to keep them, however worth evaluating in case any are felt redundant. From a physical activity perspective, face validity looks reasonable. The main variables of value could be argued to be:

- Time spent in moderate intensity PA
- Time spent in Vigorous intensity PA

From these you can derive:

- Time spent in Moderate to Vigorous PA (MVPA)

In its most simple format, this allows comparative analyses against the current Chief Medical Officer's guidelines and so has clear academic and wider value.

You can interrogate and present these data at an individual level (see Figure 1 and 2), which will allow for intra-individual variation to be explored, and for some advanced statistics that account for clustering such as multilevel modelling. This is even more valuable when you introduce a spatial/hierarchical component to the dataset.

In Figure 1a and B, the bar chart is a simple reflection of time spent in MVPA for each day of the dataset (Figure 1) and a stacked bar chart (Figure 2) showing how this participant accumulated their total volume of MVPA across the week. This user met their recommended weekly level of MVPA on their 3rd day and had a mean of 45 mins/day of MVPA. With this type of data you can explore ***how*** a person accumulates their physical activity. This participant had an elevated 1st day of which, in general, reduced linearly as they progressed through the week. Total volume, mean volume per day, and how MVPA is accumulated (e.g., the weekend warrior) can/are used as predictors of other physical and mental health outcomes.

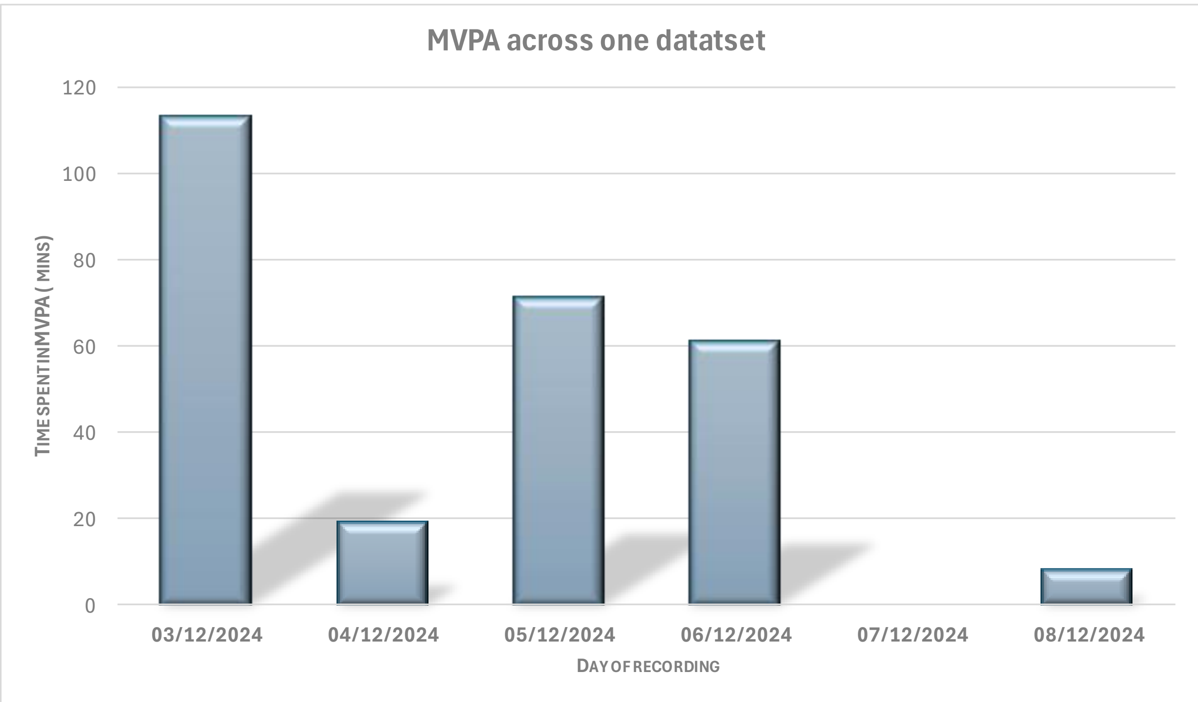

FIGURE 1: DAILY MVPA FOR ONE PARTICIPANT ACROSS THE DATASET

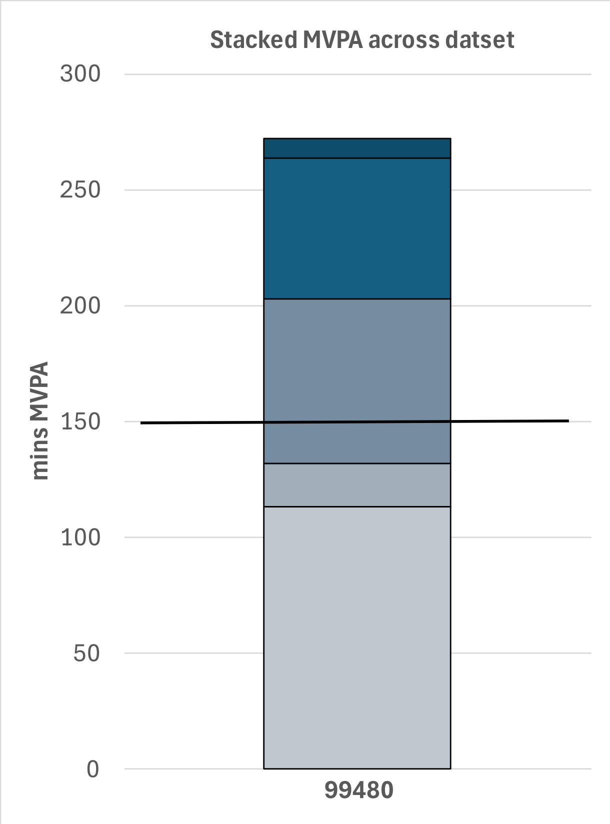

FIGURE 2: STACKED MVPA AGAINST CMO GUIDELINE FOR ADULT'S TOTAL WEEKLY VOLUME – USER 99480

Data can be aggregated and presented/explored across samples/populations at ease (see Figure 3, 4, and 5), and could easily include demographic data to further explore (e.g., Gender, SES, Age). Where Figures 1 and 2 present MVPA at an individual level, Figure 3 demonstrates this across the sample. It is easy to see that you can further derive potential variables such as (meet/does not meet guidelines), where only 1 participant would have passed the 150 minutes of MVPA/week – although 99481 was close. Further derivation of variables could be valuable for specific types of analyses, for example whether meeting the PA guidelines was a strong predictor of health outcomes such as specific cancers. You could easy run sensitivity analyses around where the threshold was (140mins instead of 150mins for instance) and demonstrates the value of this type of outcome from the Garmin device.

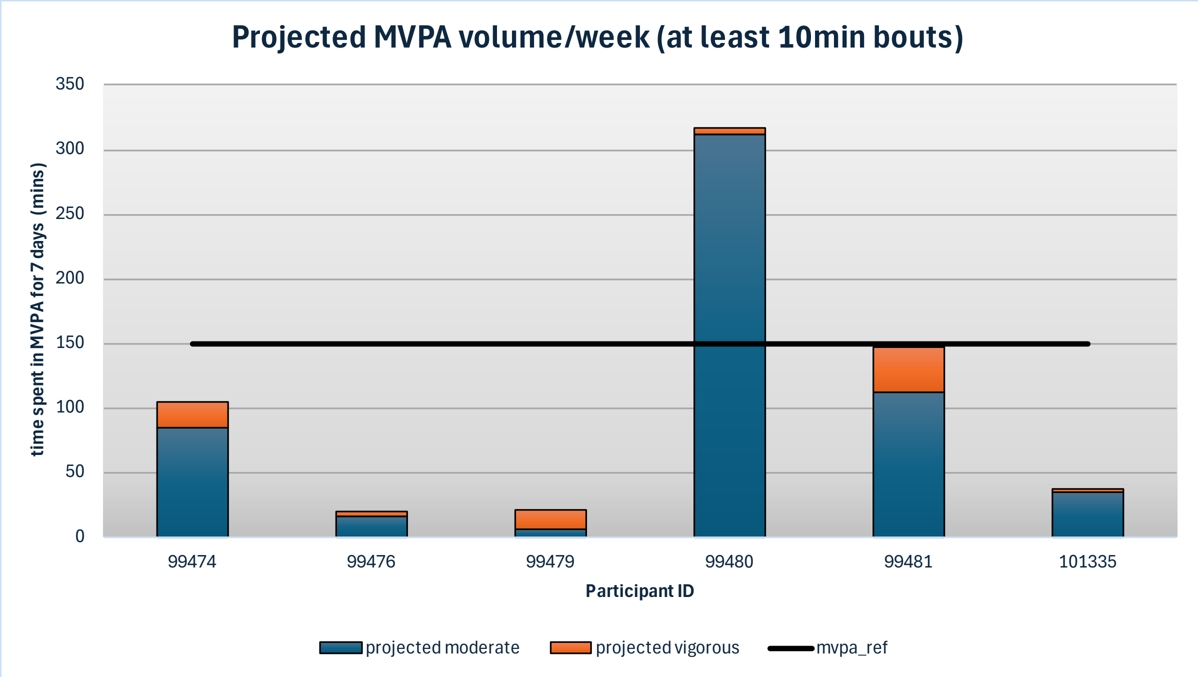

FIGURE 3: MVPA VOLUME ACROSS 7 DAYS FOR THE SAMPLE

In addition to MVPA, the standard outcome variables available from the dashboard include total time spent being active (Figure 4), a variable that is calculated as the sum of moderate minutes plus twice the vigorous minutes. A limitation of the standard variables extracted by Avicenna is that it does not include a discrete light intensity activity variable – and our current understanding of PA and health is that engaging in light activity is also extremely important for our health.

Figure 5 presents a further useful activity variable, total daily steps, across the sample. Much like MVPA, step count is a long-standing PA outcome that provides unique insight and has long been used as a simple public health message (10,000 steps/day). With the device being wrist-worn the device will use some pattern recognition from the acceleration signal to identify steps. Again, you can derive and manipulate this variable to see how these are accumulated (e.g. where and when during the day) and so has important value for academic research.

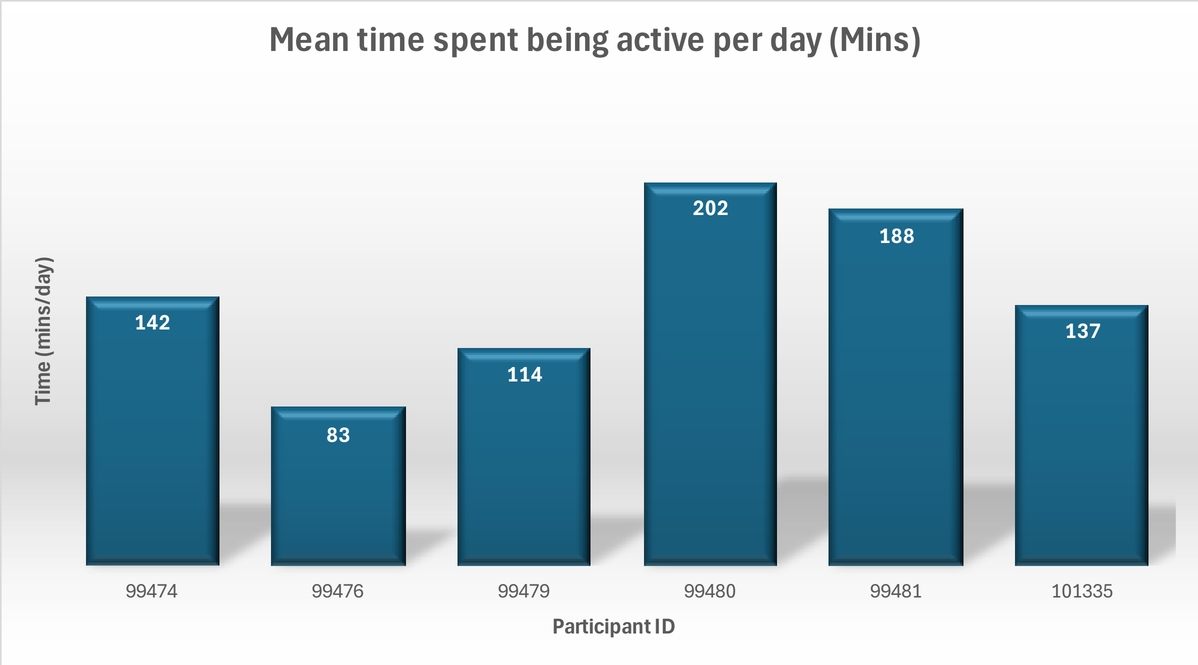

FIGURE 4: MEAN TIME SPENT BEING ACTIVE PER DAY (MINS)

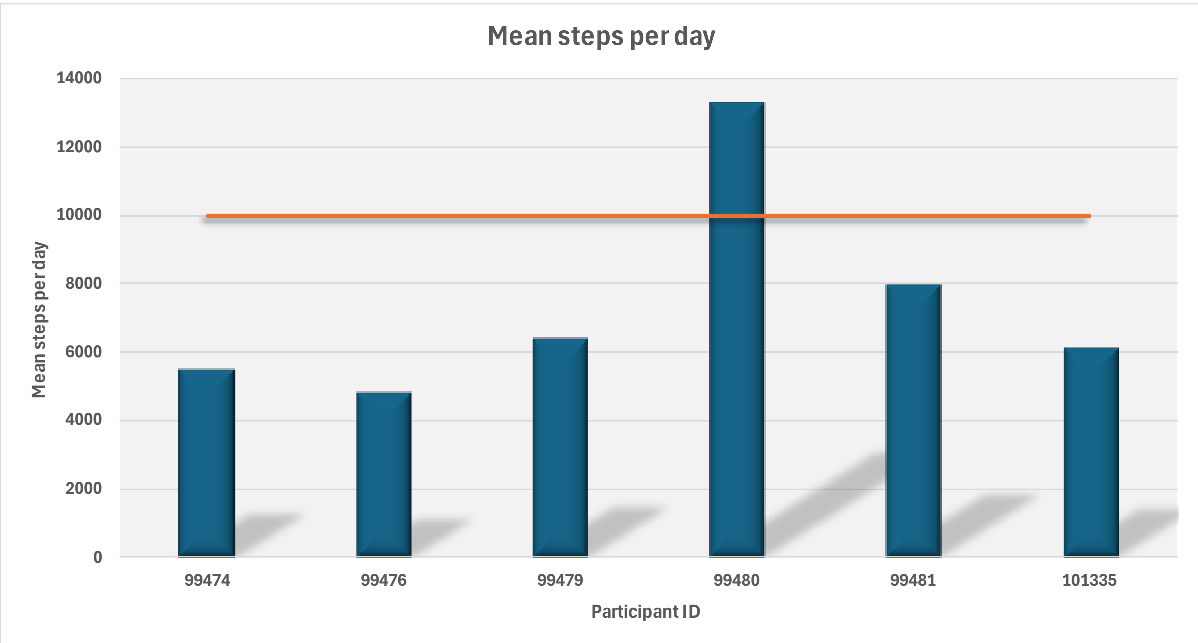

FIGURE 5: MEAN STEPS PER DAY ACROSS THE SAMPLE

15-MINUTE EPOCH DATA

Based on our assessment of the 15-minute level data it was decided there would be no additional value in presenting this within the report. From a research perspective, it does not hold any additional insight that is presented at the daily level, nor if the raw accelerometry data can be collected.

PHYSICAL ACTIVITY DATA – LIMITATIONS AND RECOMMENDATIONS

At this stage, the solution provided by Avicenna doesn't meet the requirements of the research but as far as we are aware, most – if not all – issues can be resolved and would not represent a red traffic light, if we were to use progression metrics commonly seen in evaluation design. Most would be amber lights, and Avicenna need to provide clarification on whether these can be integrated. These are presented in detail below:

Raw acceleration data – Using raw accelerometry data to measure physical activity levels offers several advantages. First, it provides highly granular and precise information about movement patterns, capturing data at a much higher resolution than pre-processed or summary data (often proprietary). This allows for the detection of subtle activity variations, such as posture changes, sedentary behaviours, or brief bursts of movement, which might otherwise be missed. Additionally, raw data enables flexibility in applying customized algorithms and thresholds tailored to specific populations or research objectives, improving the accuracy of activity classification beyond the immediate project goals. It also ensures transparency and reproducibility, as researchers can reanalyse the data using updated techniques as methodologies evolve. The raw data can be processed in open applications such as GGIR, allowing calibration/agreement studies to be conducted comparing the Garmin processed data with that of something like GGIR. The flexibility of raw data means you can explore the data in a variety of ways, extracting data for commute periods, the weekend, leisure time, and the workday. These then increase the utility of the data by having greater application to a wide range of potential users. Furthermore, raw accelerometry data can facilitate advanced analyses, such as machine learning applications, to identify complex patterns or predict health outcomes. If future research studies or external teams wanted to use the data, then anonymised raw data could be deposited in a repository for further exploration.

Sedentary time as a daily outcome – Sedentary time is the amount of time spent sitting, reclining, or lying down while expending very little energy. It is different from 'physical inactivity', which is the absence of physical activity. One can spend several hours sedentary yet still do enough physical activity to meet the recommended level of MVPA. It is an important component of the 24 movement behaviours, and independent predictor of health and wellbeing. At this stage it is not clear if time spent sedentary is explicitly classified by the Garmin device and has been queried with Avicenna. To extract the underlying data from Garmin you would need to use acceleration data and or the heart rate data. However, if the raw data is provided, then this can be classified independently using other algorithms in packages such as GGIR.

Light intensity as a daily outcome – Similar to Sedentary time, the Garmin dashboard does not extract light intensity as a standard metric. Over the last decade a vast amount of research has evidenced the Importance of light intensity PA as a predictor of health and wellbeing and is a central component of 24 movement behaviours. If comparing to the way Garmin classifies MVPA, light intensity activity could similarly be extracted based on heart rate data that sits within a specific percentage of maximum heart rate (e.g., 50-70%). We have queried Avicenna on this so await their response. Again, much like sedentary time, this can be extracted via open processing packages such as GGIR.

Wear time – Although the envisaged protocol is one where participants will be asked to wear the smartwatch for 24 hours, it would be valuable to have some indication if the watch has been removed. Non-wear time is categorised by zero/very little acceleration being recorded by the device. However, as sitting and quiet standing can also record as zero or little acceleration (and low heart rate) there is a requirement to try and differentiate so as to not misclassify non-wear as sedentary time and vice versa. Additionally, where there is some requirement to have a specific number of valid hours per day before being included in any analysis, an objective measure of non-wear time can be made about compliance to the protocol. You can, and should, ask participants to record any time the device has been removed however the heart rate data, accelerometer data, skin temperature, and light sensors can be used to evaluate this. It is our understanding that this can be derived from the Garmin SDK. Like many of the other variables, the raw acceleration data can also be processed, independently, to evaluate non-wear time – further increasing the value of ensuring its extraction.

Outcomes variables at a higher frequency (1-minute epochs) – Where Garmin processed Daily summary variables are valuable for research purposes and can be used in a variety of different contexts, e.g., as predictors, mediators, covariates and confounding, and outcomes within cross-sectional and longitudinal analyses, the opportunity to derive the same variables at a higher frequency (e.g., 1-minute intervals) opens up the potential to integrate more refined temporal analyses (e.g., focusing on the commute, the work day, or leisure time). Moreover, the further benefit of higher frequency health data is the opportunity to link to high frequency spatial data (the GPS). This transforms the value of the wider investment as the dataset can then be used to answer important questions about how space, place, and time determine health and wellbeing. Where do people go (and not) and why? Does where you live, or work influence your activity levels? Does engagement with greenpace matter for health and wellbeing. How can GPS and accelerometry be used to derive travel mode? See the final section of the report for some more detailed examples. Variables we suggest would be important to extract include Heart rate, Light intensity PA, Moderate Intensity PA, Vigorous Intensity PA, Sedentary time

SLEEP – DAILY DATA

We recognise that the Understanding Society research team has specific expertise in sleep and so understand that they might be doing more detail analyses and have more detailed feedback on the value of the device for measuring sleep. However, the Garmin device collects and can easily present daily level sleep metrics, and Figure 6 captures the mean duration of sleep within each component of the sleep cycle, presented as a stacked bar graph to show total sleep duration. You can further derive ratios of each to the total and the standard outcomes also allow for some sleep quality indicators (not shown).

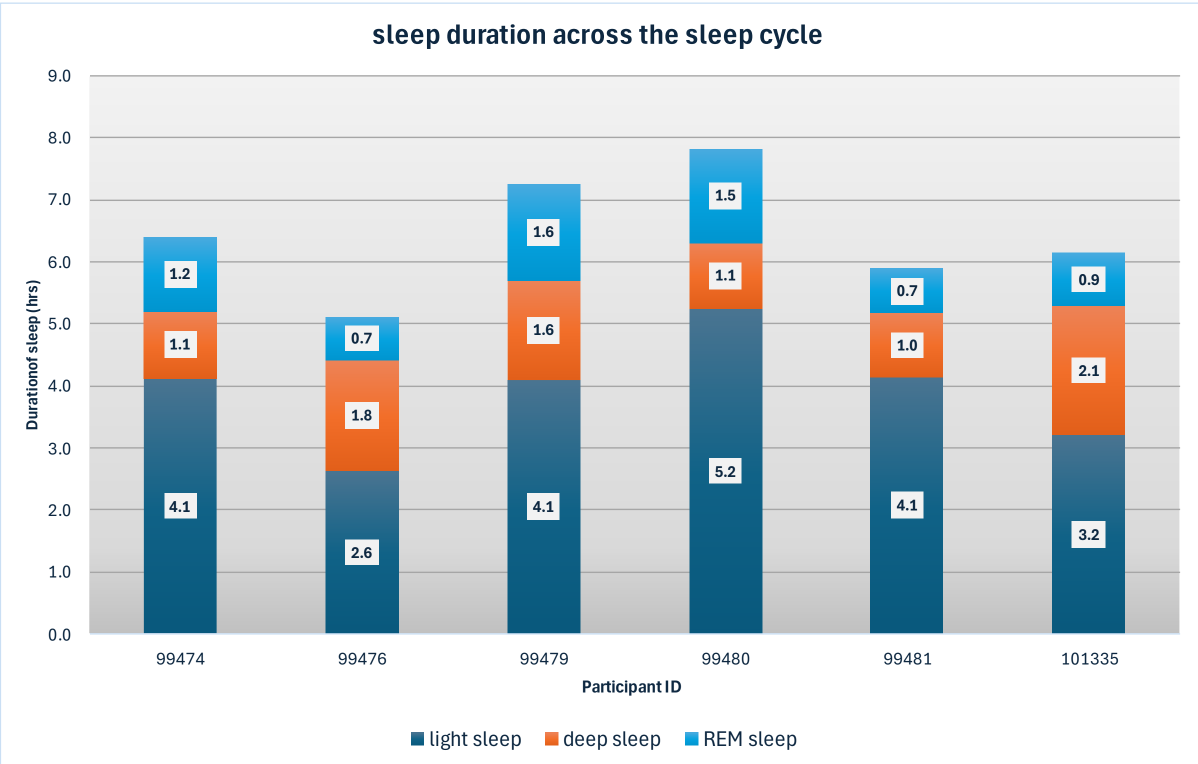

FIGURE 6: MEAN SLEEP DURATION ACROSS EACH COMPONENT OF THE SLEEP CYCLE

GPS Data

A total of 510,767 GPS locations were recorded across 6 users. Of these, 78,193 locations (15.3%) were duplicates at 1-second rounded satellite timestamps. Duplicates varied by operating system (OS) and user (Table 1). Android devices typically had lower percentage of duplicates (6.7-16.6%) than iOS devices (23.6-61.2%). Android devices also had much higher total GPS readings (47,774-340,397) than iOS (129-9,672).

Reasons for lower GPS count of iOS is covered in 'Problem GPS'. High count of raw GPS from Android devices is due to higher sampling frequency, covered in 'GPS Sampling Frequency'.

Data should be cleaned to remove duplicates based on 1-second rounded satellite timestamps, preserving the location with the lowest accuracy for each timestamp. Having such high-frequency raw data to work from does increase the positional accuracy after cleaning, however, so is more advantageous than not.

Table 1: Number of GPS and duplicates by user and operating system (OS)

OS	User	Total GPS	Unique GPS	Duplicate GPS	Percentage duplicates
Android	99472	47,774	44,594	3,180	6.7
Android	99479	109,607	97,258	12,349	11.3
iOS	99481	9,672	4,418	5,254	54.3
Android	99476	340,397	283,818	56,579	16.6
iOS	99474	3,188	2,436	752	23.6
iOS	99480	129	50	79	61.2
Both	All	510,767	432,574	78,193	15.3

PROBLEM GPS

Two users (99474 and 99480) recorded very few GPS collected compared to other users. However, they do have health data, suggesting the Avicenna part of the linkage is running nominally, or there wouldn't be any data from the user at all. Similarly, this was not a GPS signal error, as Garmin Connect retained GPS tracks recorded as activities (e.g. hike) for user 99480.

Common denominator: Both users have older version of iPhone (11 pro; 12 pro).

Issue: The Garmin app on iPhones may not syncing the data continuously with the Garmin servers. This can happen either due to user 'hard closing' the Avicenna app by inadvertently 'swiping away' the app to close, or the enabled battery optimisations on the phone, or an unstable network connection that results in data loss.

Solution: Instruct users to keep Avicenna app open throughout test period (avoid swiping away). Make sure the battery optimisations are off, or the Garmin app is exempt. Older iPhones may have battery optimisations turned on by default.

GPS Sampling Frequency

Before data are cleaned (one second-level duplicates removed) there were significant platform-specific behaviour in GPS data collection. Android users (users: 99472, 99479, 99476) show frequent high-sampling periods (>= 30 GPS points per 30 second epoch) and can reach very high sampling rates (up to 126 points per 30 second epoch, or ~4Hz). This was not the case with iOS users iOS users (user_id: 99480, 99481, 99474), who never recorded >30 points per 30 second epoch. This suggests iOS devices are either: 1. Collecting data at a lower fixed rate; 2. Applying some smoothing/filtering.

As noted, high-frequency sampling has advantages, in terms of flexibility to retain the most accurate GPS positions per epoch. However, this platform difference could be important for:

1. Analysis methodologies
2. Battery life implications
3. Storage requirements

Epoch-level Data Coverage

Removal of duplicate GPS left 432,574 locations independent to 1-second rounded satellite time. However, 1-second intervals are too frequent to be useful and therefore need to be aggregated to 1-minute epochs, which can be matched with 1-minute PA or HR epochs.

Using cleaned data (duplicates removed) matched against the start and end wear dates provided in user reports, we can see variation in epoch coverage by Users during the test period (Table 2).

Table 2: Expected and observed number of 60-second epochs with data

OS	User	Start date	End date	Expected epochs	Observed epochs	Epoch coverage (%)
Android	99472	01/12/2024	09/12/2024	11,520	1,601	13.9
iOS	99474	30/11/2024	06/12/2024	8,640	1,169	13.5
Android	99476	30/11/2024	04/12/2024	5,760	2,826	49.1
Android	99479	26/11/2024	29/11/2024	4,320	691	16.0
iOS	99480	03/12/2024	10/12/2024	10,080	24	0.2
iOS	99481	27/11/2024	10/12/2024	18,720	2,850	15.2

Temporal Patterns in Epoch Coverage

When 1-minute epoch coverage was examined in 3-hour blocks (Figure 7), some clear patterns emerged:

- User 99481 (iOS) showed the most consistent with steady, but lower, coverage around 12-16% and very regular pattern suggesting fixed sampling rate.
- User 99476 (Android) showed the highest coverage, with high consistency (~36-42% throughout).
- User 99472 (Android) had very variable coverage, which was better in early morning blocks (00:00-06:00) with poor coverage in afternoon.
- User 99479 (Android) had higher coverage during daytime with best coverage 12:00-15:00 (block 5) but poor overnight coverage suggesting device removal overnight.

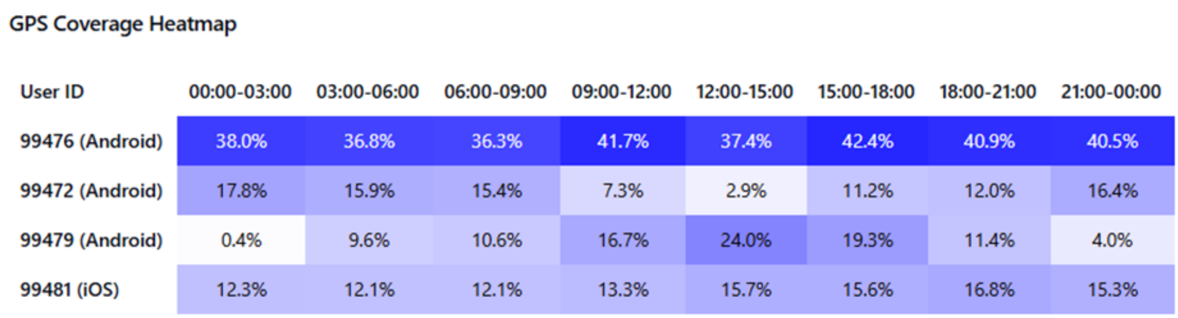

Figure 7: Heatmap of temporal patterns in 1-minute epoch coverage across the day (3-hour blocks) for each user.

Protocol Test Data Processing and Analysis Methods

This section describes how three distinct protocols were tested:

- Protocol 1: Smartphone Absent, Smartwatch Present
- Protocol 2: Smartphone Present, Smartwatch Absent
- Protocol 3: Both Devices Present

GPS data were processed in sequential steps:

- First, the most accurate GPS location within each 60-second epoch identified using minimum accuracy value
- Next, locations within protocol tests time bounds were classified into journey types (outward, wait, inside, return) based on timestamp comparisons with trip timing data provided by participants:
    - ‘Wait’ journey classifications accounted for outdoor static testing of participants who waited as instructed for 5-mins outside destinations
    - ‘Inside’ journey classifications accounted for indoor static testing of participants who entered destinations for at least 5-mins
    - ‘Outward’ and ‘return’ journey classifications represent dynamic/moving journey segments when participants walked to and from the destination
- GPS data collection performance across protocols, journey types, and participants was assessed through three key metrics that were compared:
    - GPS Completeness: Calculated as percentage of observed vs expected GPS points (60-second epochs)
    - GPS Positional Uncertainty: Mean (SE) uncertainty measured as radius (meters) of likely true location. Lower values indicate higher certainty of location’s true position.
    - GPS Accuracy: Calculated as the percentage of GPS points within different accuracy bands ("Under 10m," "10 to 20m," "20 to 50m," and "Over 50m")
- Tracking performance across protocols, journey types, and participants was assessed through two key metrics that were compared:
    - Route adherence: Distance between GPS points and known travel path taken during outward/return journeys supplied as maps by participants). An example route visualisation is show in Figure 8.
    - Location adherence: Distance between GPS points and true location of destination during stationary periods

Results are organised into static (wait, inside) and dynamic (outward, return) journey segments to evaluate protocol performance under different tracking contexts/conditions.

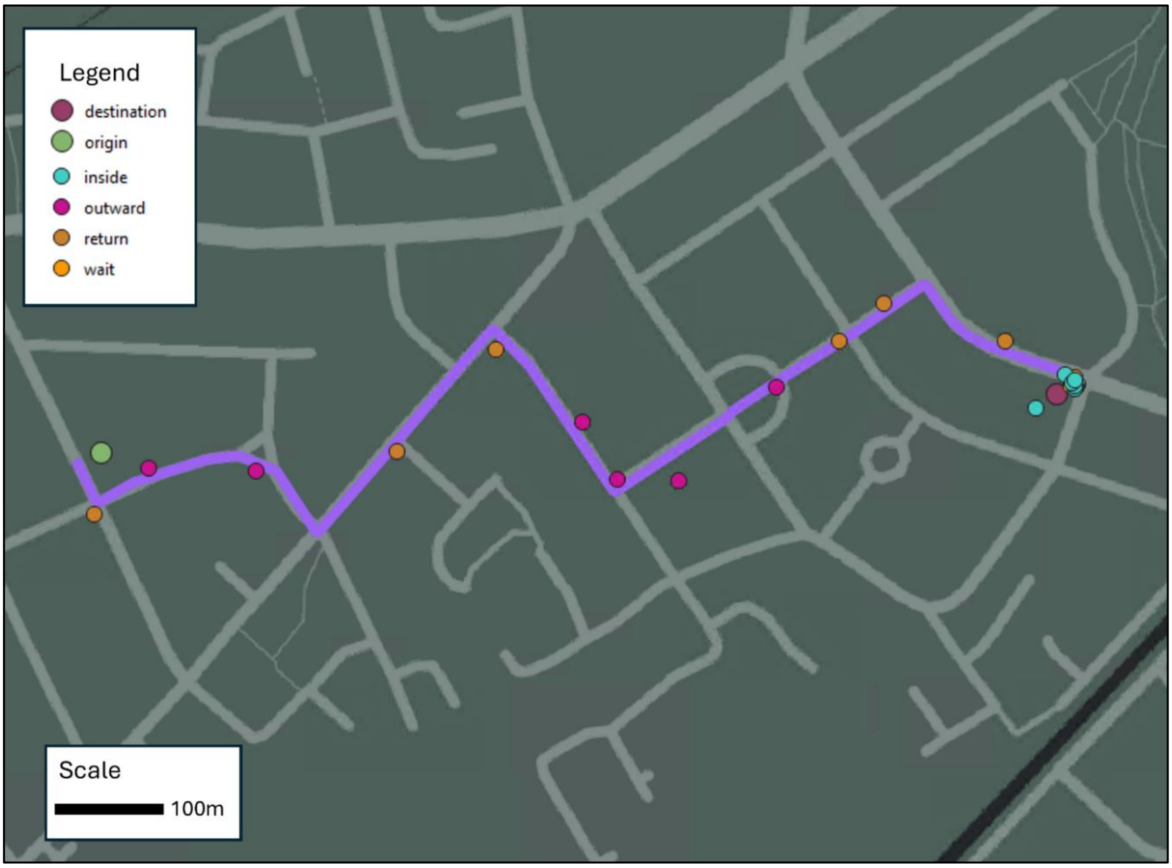

Figure 8: Example of GPS locations between origins and destinations classed as outward and return journeys, along with waiting static outdoor test, and static indoor test. The solid line indicates the true route path taken by the user. Note that 'wait' points are hidden by overlapping "inside" points.

PROTOCOL DATA AVAILABILITY

The start and end date/times each participant recorded each of the three test protocols was matched against GPS observations to determine how much data were available to use in protocols assessments. A summary of users (and phone operating systems) with GPS data for each Protocol is shown in Table 3. This indicates the sample sizes on which subsequent protocol assessments are based are variable, with larger sample for testing Protocol 2 (Smartphone Absent, Smartwatch Present) and lowest sample for Protocol 3 (Both Devices Present). This means that findings from assessments should be taken as preliminary. Issues around GPS data collection from two further iOS users, who had no data coverage within any protocol period, were given in the initial report. It is likely that certainty around conclusions of assessments will be strengthened with a second trail in which potential contributory factors have been mitigated.

All subsequent analyses were conduct for participants with any GPS data collected during protocols. However, this does include some who have lower epoch coverage (<30% of expected epochs).

Table 3: Number of participants (and operating systems) with sufficient data to evaluate trial protocols.

Protocol	Description	Mean completeness (%)	Number participants	Number iOS devices	Number Android devices
1	Smartphone Absent, Smartwatch Present	65.4	4	1	3
2	Smartphone Present, Smartwatch Absent	74.7	3	1	2
3	Both Devices Present	59.1	2	0	2

Indoor vs. Outdoor Static Testing

The ability to reliably detect outdoor stationary periods, at dwell points, and indoor periods when a building has been entered, are key for object exposure assessment. These are quantified here as 'wait' and 'inside' journey types.

Static Completeness

The high variability in wait period completeness (standard errors of 61%, 47%, and 50% for Protocols 1-3 respectively) prevents definitive conclusions about outdoor stationary detection reliability (Figure 9). While Protocol 2's 94% mean completeness suggests strong performance, the wide confidence intervals indicate inconsistent detection across participants.

Indoor completeness was consistently lower (54-69%) with smaller standard errors (13-27%), suggesting more reliable but systematically reduced performance across protocols. The similar performance across protocols, including Protocol 3's 54% (±13%) completion rate, indicates indoor signal degradation is a fundamental GPS limitation rather than a device-specific issue.

The outward-return journey pattern shows increasing completeness across all protocols, most pronounced in Protocol 3 (56% to 76.5%). However, this trend's reliability is limited by:

- Small sample sizes (n=2-4 per protocol)
- High variability (standard errors 9-22%)
- Potential confounding factors (weather, building materials, satellite positions)

Further trials should include repeated measures and controlled environmental conditions to better isolate device performance from external variables.

Static Positional Uncertainty

Positional uncertainty when devices were indoors (Figure 10) varied substantially (15-24m), with Protocol 2 showing poorest performance (24.29m ± 8.76m). This inverse relationship with its high completeness rate (62-94%) suggests smartphones may prioritise maintaining connections over positional certainty indoors.

Positional uncertainty of wait periods was most consistent (8.7-11.1m, SE 3.9-5.0m across protocols), indicating reliable static positioning regardless of device configuration. However, large standard errors, particularly in Protocol 3 (±5.0m), suggest environmental factors may significantly influence stationary positional certainty.

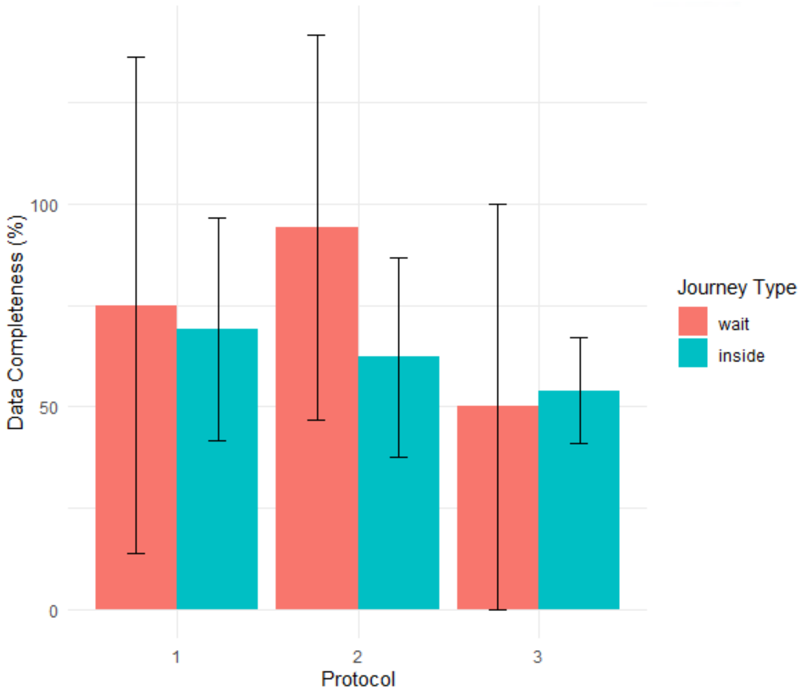

Figure 9: Mean (SE) data completeness (percentage of expected GPS points observed at 60-second intervals) across static journey segments (wait; inside) for each protocol: Protocol 1: Smartphone Absent/Watch Present (n=4); Protocol 2: Smartphone Present/Watch Absent (n=3); Protocol 3: Both Present (n=2).

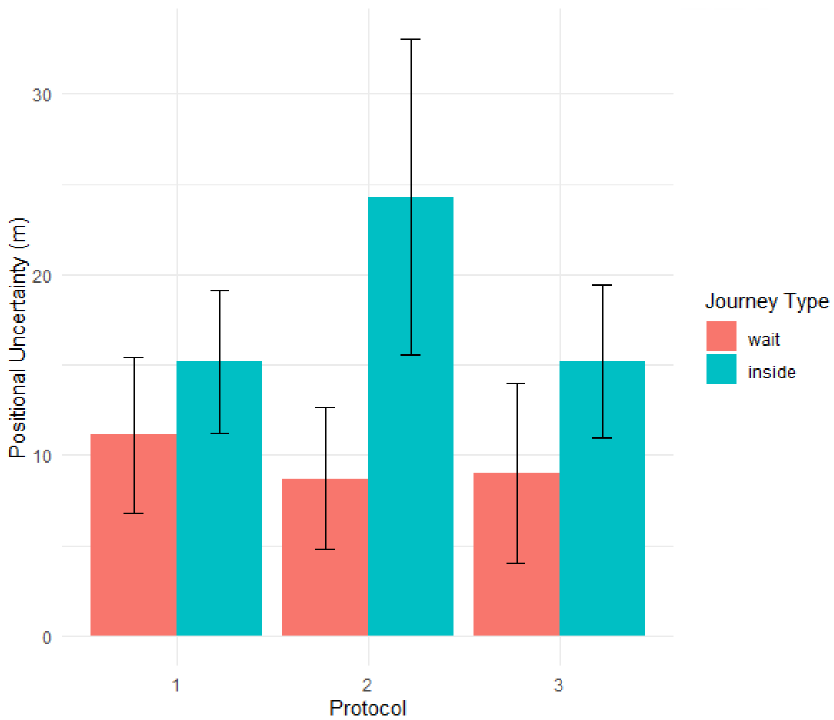

Figure 10: Mean (SE) GPS uncertainty (meters) across static journey segments (wait; inside) for each protocol: Protocol 1: Smartphone Absent/Watch Present (n=4); Protocol 2: Smartphone Present/Watch Absent (n=3); Protocol 3: Both Present (n=2). Lower values indicate higher certainty, and therefore greater accuracy, of location's true position.

STATIC GPS ACCURACY

GPS accuracy showed distinct patterns between indoor and outdoor static periods (Figure 11). During outdoor waits, all protocols maintained good precision, with 50-67% of fixes achieving high quality (<10m uncertainty) and remaining fixes mostly within moderate accuracy (10-20m). Protocol 2 (smartphone) performed best outdoors, with 67% of fixes under 10m.

Indoor segments showed markedly degraded accuracy, which is to be expected. Protocol 1 (watch) maintained some high-quality fixes (25% under 10m), but Protocol 2 showed substantial degradation with 67% of indoor fixes in the 20-50m range. Protocol 3 (both devices) showed consistent but moderate indoor accuracy (100% of fixes 10-20m).

This indoor-outdoor contrast suggests building interference affects all device configurations, though smartphones (Protocol 2) show the starkest performance decline when moving indoors. The complete absence of very poor-quality fixes (>50m uncertainty) indicates basic position reliability is maintained across all conditions.

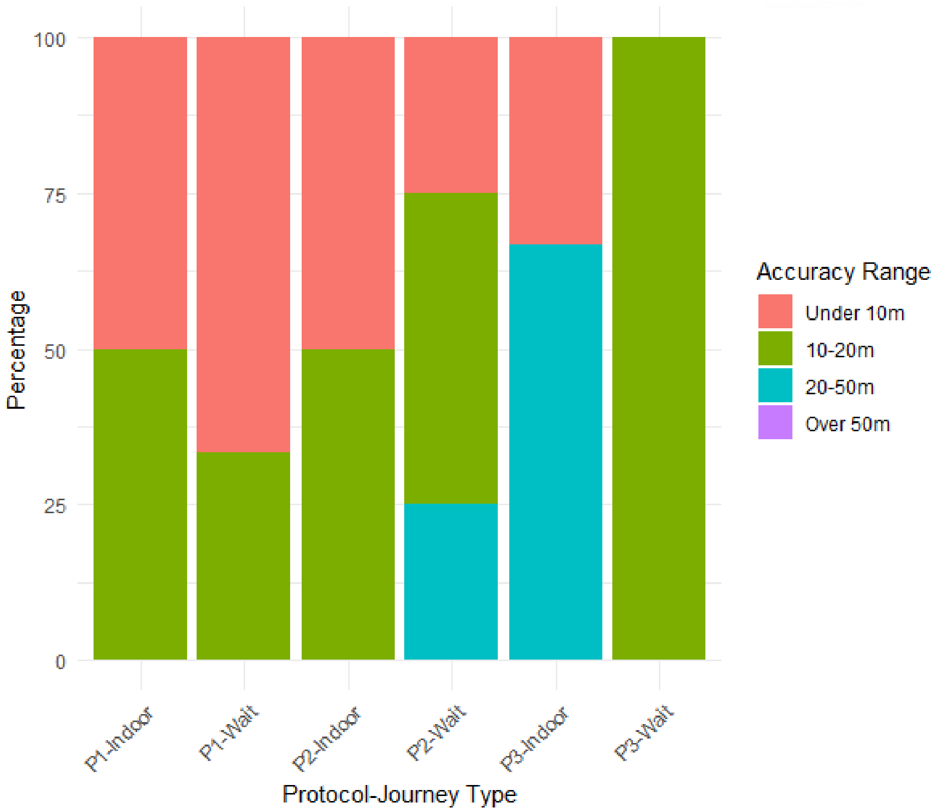

Figure 11: Percentage of GPS points within different accuracy bands ("Under 10m," "10 to 20m," "20 to 50m," and "Over 50m") for static journey types (wait, inside) by protocol: Protocol 1: Smartphone Absent/Watch Present (n=4); Protocol 2: Smartphone Present/Watch Absent (n=3); Protocol 3: Both Present (n=2).

Static Location Adherence

Analysis of aggregated GPS point distances reveals distinct patterns across protocols for static locations (Figure 12), with a clear hierarchy emerging across protocols. Protocol 3 achieved the most precise positioning, with consistent performance between wait (8.3m ± 6.3m) and inside (9.5m ± 7.0m) periods. Protocol 2 maintained moderate precision with similar indoor-outdoor performance (wait: 21.0m ± 7.0m; inside: 21.4m ± 10.9m.

Protocol 1 (watch) showed substantial positioning errors, with large mean distances and high variability (wait: 361.4m ± 348.1m; inside: 465.8m ± 198.5m). This marked degradation suggests fundamental limitations in smartwatch-only positioning during stationary periods, contrasting sharply with the stability of dual-device and phone-only configurations.

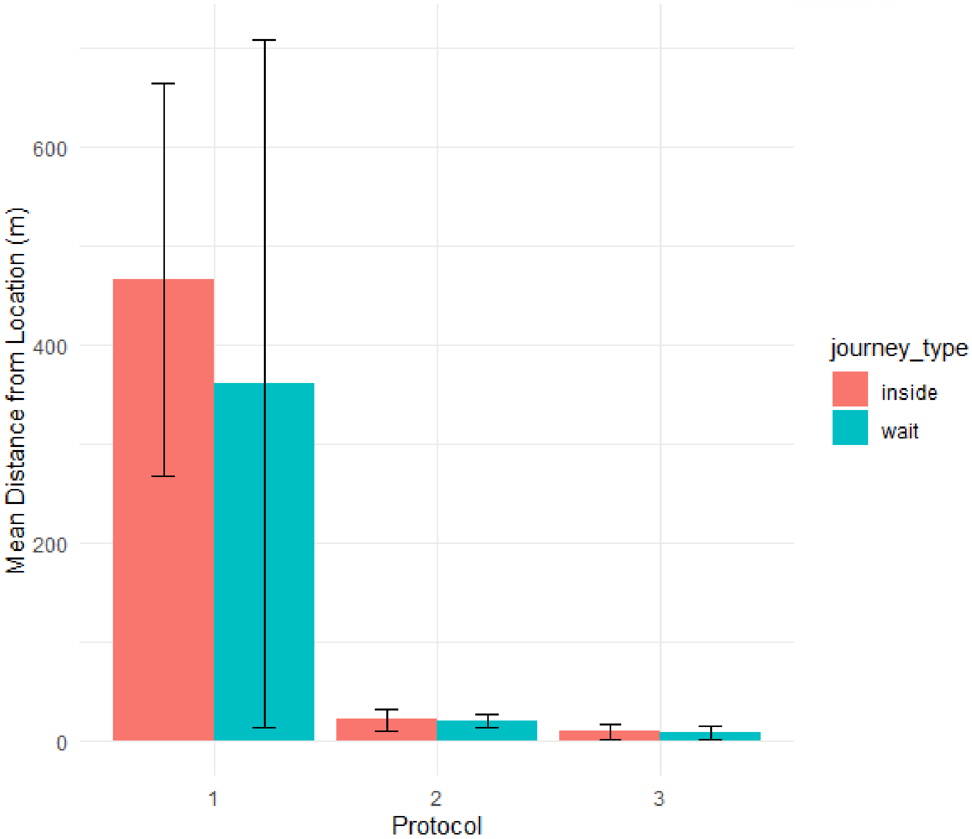

Figure 12: Mean (SE) location adherence, indicating the distance between indoor/outdoor static journey segments (inside, wait) and the known destination location.

Outward vs. Return Dynamic Testing

Journey segments (outward and return trips) provide insights into GPS performance during continuous movement - a critical aspect for capturing travel patterns and route choices in exposure assessment. Evaluating both journey types allows assessment of GPS initialisation effects (outward) and sustained/improved tracking capability (return).

DYNAMIC COMPLETENESS

Completeness during moving journey segments showed systematic variation across protocols (Figure 13). Protocol 2 maintained relatively high completeness for both outward (67.7% ± 21.9%) and return journeys (74.3% ± 16.6%), suggesting smartphones provide consistent tracking during movement. In contrast, Protocol 1's lower completeness rates (outward: 57.0% ± 11.8%; return: 65.0% ± 14.9%) indicate smartwatches may struggle with continuous position logging during extended movement.

Protocol 3's pattern was particularly notable, showing the lowest outward journey completeness (56.0% ± 13.0%) but strong return journey performance (76.5% ± 9.5%). This improvement in completeness parallels the accuracy stabilization seen with dual devices, though the small sample size (n=2) limits definitive conclusions. The consistent trend toward improved completeness during return journeys across all protocols suggests GPS performance benefits from sustained outdoor exposure and established satellite connections.

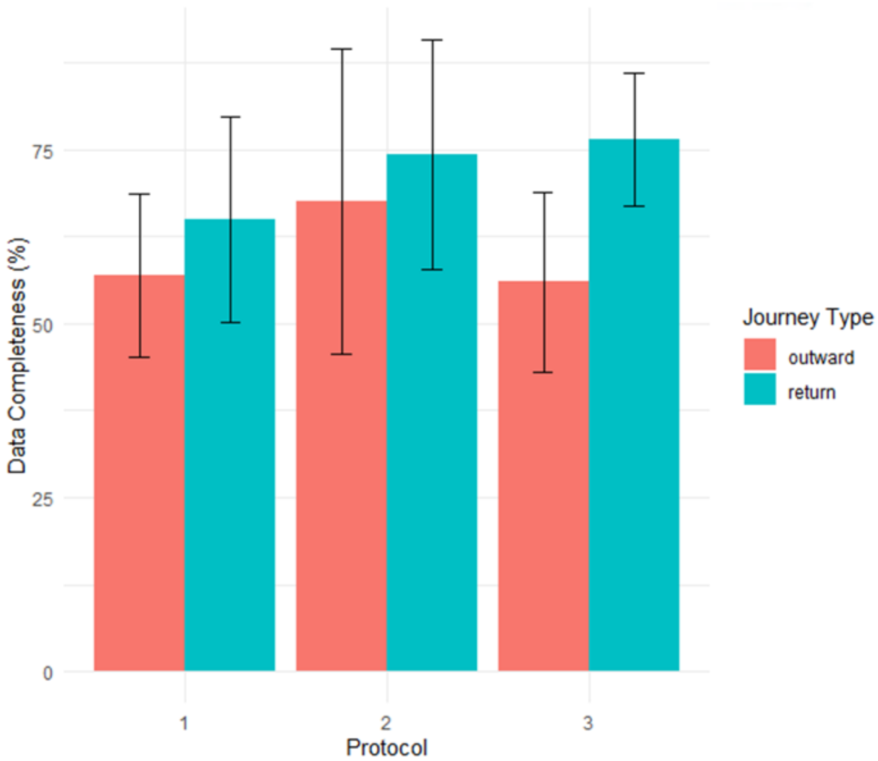

Figure 13: Mean (SE) data completeness (percentage of expected GPS points observed at 60-second intervals) across dynamic journey segments (outward; return) for each protocol: Protocol 1: Smartphone Absent/Watch Present (n=4); Protocol 2: Smartphone Present/Watch Absent (n=3); Protocol 3: Both Present (n=2).

Dynamic Positional Uncertainty

Journey segment accuracy showed distinct patterns across protocols (Figure 14). While Protocol 3 maintained consistent certainty between outward and return journeys (10.55m ± 4.33m to 9.28m ± 1.90m), both single-device protocols showed notable degradation in uncertainty, Protocol 1's positional certainty decline from outward (10.11m ± 3.41m) to return journeys (18.71m ± 5.44m) was particularly pronounced, suggesting smartwatches may struggle to maintain accuracy over journey duration. The reduced standard error in Protocol 3's return journeys points to potential stabilization benefits of dual devices, though the limited sample size (n=2) necessitates cautious interpretation.

This journey-based variation, combined with the indoor/outdoor findings, suggests device configuration impacts both initial GPS acquisition and accuracy (certainty) maintenance throughout movement periods. The data hints at complementary device strengths, with smartphones excelling at connection maintenance (higher completeness) and dual devices potentially offering more stable accuracy, though larger trials are needed to confirm these patterns.

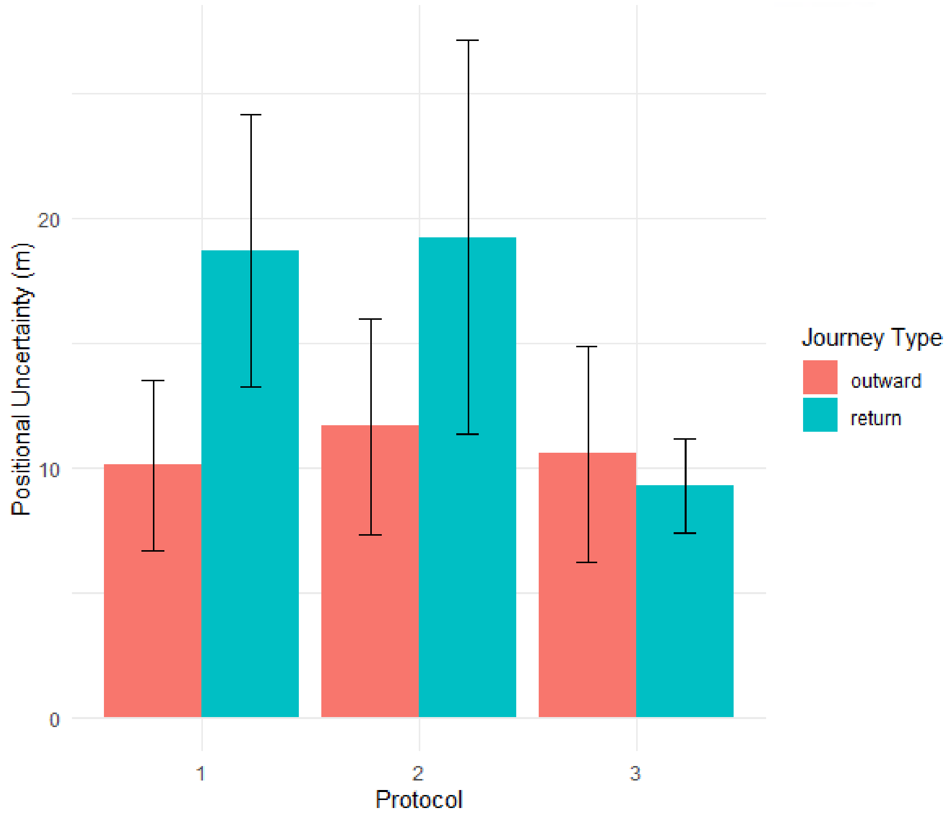

Figure 14: Mean (SE) GPS uncertainty (meters) across dynamic journey segments (outward; return) for each protocol: Protocol 1: Smartphone Absent/Watch Present (n=4); Protocol 2: Smartphone Present/Watch Absent (n=3); Protocol 3: Both Present (n=2). Lower values indicate higher certainty, and therefore greater accuracy, of location's true position.

Dynamic GPS Accuracy

GPS position uncertainty during movement revealed distinct protocol performance patterns (Figure 15). During outward journeys, all protocols maintained good precision with 33-50% of fixes achieving high quality (<10m) and most remaining fixes within moderate uncertainty (10-20m). Protocol 3 (both devices) showed most consistent performance, maintaining this distribution for return journeys.

Single-device protocols showed notable degradation during return journeys. Protocol 1 (watch) declined to 25% high-quality fixes with 50% in the 20-50m range. Similarly, Protocol 2 (phone) showed increased uncertainty, with 67% of return journey fixes in the 20-50m range. This suggests potential signal degradation over journey duration when using single devices.

The consistent performance of Protocol 3 across both journey segments indicates dual devices may help maintain position precision throughout movement periods, though the limited sample size warrants cautious interpretation.

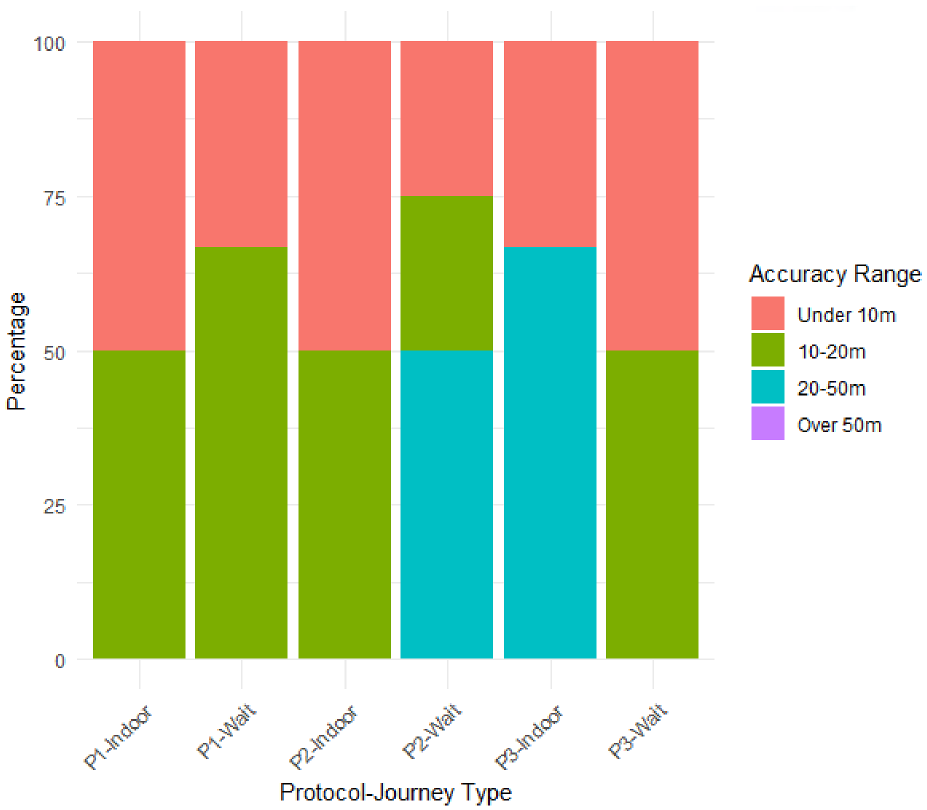

Figure 15: Percentage of GPS points within different accuracy bands ("Under 10m," "10 to 20m," "20 to 50m," and "Over 50m") for dynamic journey types (outward, return) by protocol: Protocol 1: Smartphone Absent/Watch Present (n=4); Protocol 2: Smartphone Present/Watch Absent (n=3); Protocol 3: Both Present (n=2).

Dynamic Route Adherence

Analysis of route adherence during movement tracking revealed distinct patterns in location error and consistency (Figure 16), Protocol 3 (Both Devices Present) superior route tracking precision, with minimal deviation from mapped paths (outward: 8.2m ± 0.7m; return: 6.1m ± 0.3m) and notably small standard errors indicating consistent performance.

Protocol 1 (Watch present) showed moderate, stable tracking (outward: 13.8m ± 3.7m; return: 10.9m ± 2.2m), while Protocol 2 (phone) exhibited increasing deviation over journey duration (outward: 15.1m ± 5.1m; return: 19.4m ± 10.0m). This pattern suggests potential signal drift with phone-only tracking, contrasting with the stability of watch-based positioning during movement.

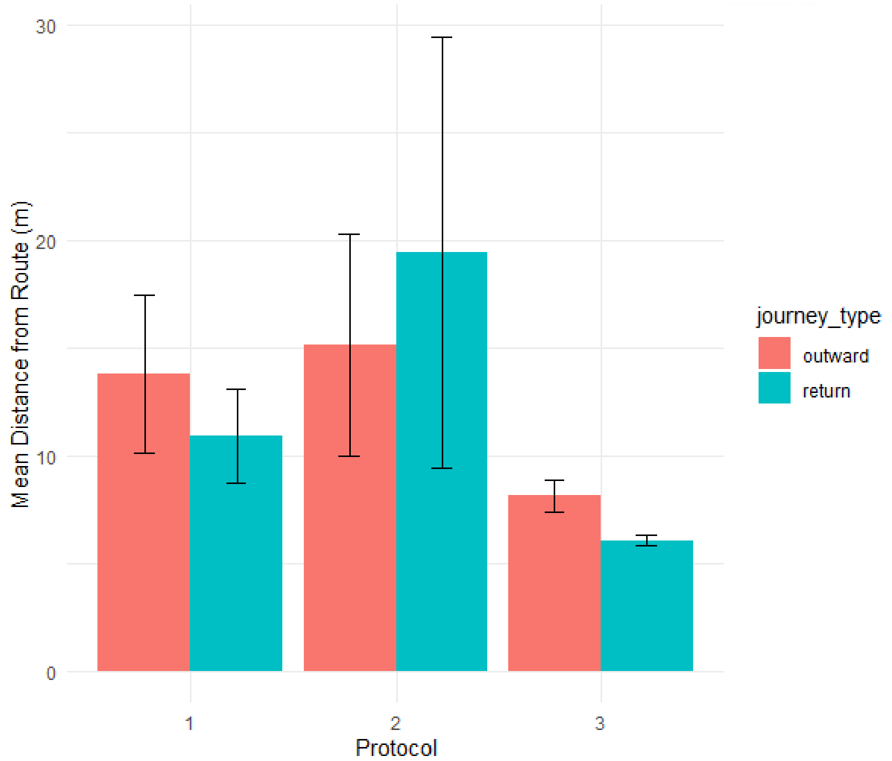

Figure 16: Mean (SE) route adherence, indicating the distance between active journey segments (outward, return) and the known route.

SUMMARY OF PROTOCOL TRIAL FINDINGS

Analysis of GPS tracking revealed Protocol 3 (dual devices) offers the most balanced solution for studies requiring both health metrics and location data. While not maximizing completeness, it achieved best route adherence (6-8m) and most consistent static positioning (8-9m), suggesting reliable spatial tracking alongside health data collection.

Protocol 1 (smartwatch) demonstrated adequate movement tracking but poor static positioning, indicating watch-only scenarios (when phones are forgotten) may compromise location accuracy while maintaining essential health metrics. Protocol 2's strong completeness becomes less relevant given the requirement for watch-based health monitoring.

Journey patterns revealed consistent GPS initialization effects and indoor positioning challenges across all configurations, suggesting these are fundamental limitations rather than device-specific issues.

The implementation of additional Global Navigation Satellite System (GNSS)quality metrics would substantially improve data validation and preprocessing capabilities. DOP (Dilution of Precision) metrics provide crucial satellite geometry information – essential for assessing position reliability independently of signal strength. Combined with Signal-to-Noise Ratio (SNR) data, this would enable differentiation between genuine indoor locations and poor reception areas, particularly valuable during building entry/exit transitions. These metrics would allow weighted quality assessment before epoch aggregation, strengthening journey classification confidence and environmental impact understanding.

Protocol Implementation Recommendations (Priority Ranked)

Essential Actions:

- Deploy dual-device setup as standard protocol (Critical)
- Maintain strict watch compliance for health metrics (Critical)
- Implement GNSS quality metrics via server APIs (High)
  - DOP for satellite geometry assessment
  - SNR for indoor/outdoor validation
  - Weighted quality preprocessing

Study Design Optimisation:

- Include GPS initialisation buffer periods (High)
- Implement geofencing of key indoor locations with SNR validation (Medium)
- Account for journey duration in quality metrics (Medium)
- Establish signal stabilisation periods (Medium)

Technical Implementation:

- Set DOP thresholds for point filtering (High)
- Apply SNR-based indoor/outdoor transition validation (High)
- Optimise epoch aggregation (Medium)
  - Consider SNR and DOP alongside accuracy for point selection within epochs
  - Weight points based on combined quality metrics
  - Flag epochs where all points exceed quality thresholds
- Monitor satellite geometry metrics (Medium)

Known Limitations (Impact Level):

- Indoor positioning reliability (Severe)
- Initial journey quality (Moderate)
- Individual/environmental factors (Moderate)
- iOS compliance issues (Moderate)
- Watch-only positioning accuracy (Moderate)

RESEARCH VALUE OF THE SMARTWATCH DATA

From the preceding sections, it is hopefully clear that the health/physical activity and GPS data have both independent and combined value when considering potential research value. This section will provide a brief explanation of each followed by some example health/GPS research questions that can be interrogated.

PA/HEALTH DATA

Independent of the GPS, the physical activity, sleep, and sedentary variables are of equal utility. Conceivably, all can be used as predictors, mediators on pathways (often health behaviours are precursors to more direct physical and mental health outcomes), or standalone primary outcomes. More recently, scientists have taken a 24-hour movement behaviour perspective; the 24-hour period is distributed and represented by either sleep, physical activity, or sedentary time on a continuum from no movement to high movement. Consequently, researchers have now begun to explore the *combined* effects of these movement behaviours on health rather than in isolation (Rollo, Antsygina, and Tremblay, 2020). A guiding research question has been “what is the optimal combination of these behaviours for health (e.g., high PA, low SB, and high sleep), and does this explain variation in specific health outcomes (e.g., adiposity, cardiometabolic health, or mental health)?” The data can be linked at individual level much like other social, economic, behavioural, biomarker, genetic, and administrative data collected. In the context of **life course research**, device-measured data provide critical insights into how early-life behaviours influence health outcomes in later years. Researchers can identify critical periods where interventions may have the greatest impact and explore the role of sleep and activity in healthy ageing. The ability to track participants over time also allows for the examination of behavioural compensations, such as changes in physical activity patterns in response to evolving health conditions or lifestyle demands.

GPS DATA

High-frequency GPS data (e.g., 60-second, 1-minute epoch), particularly that collected at same frequency as health data to allow time-matched merging of data, allows researchers to gain detailed insights into human behaviour by capturing fine-grained movement patterns over time. This data provides a deeper understanding of spatial-temporal behaviours, such as travel routines, mobility habits, and responses to environmental stimuli. Researchers can analyse how individuals navigate urban landscapes, access services, and interact with their surroundings, offering valuable information for transportation planning, public health interventions, and social sciences.

Different scientific fields leverage high-frequency GPS data to address specific research questions. In **public health**, researchers investigate how movement patterns influence exposure to environmental risks, and social determinants of health. **Urban planning** utilises GPS data to optimise infrastructure development, assess traffic congestion, and improve public transport efficiency. Meanwhile, in **ecology**, human mobility data aids in understanding interactions with natural environments, such as green space utilisation and human-wildlife conflicts.

In the field of **social sciences**, GPS data helps answer questions about social interactions, community engagement, and spatial segregation. Researchers can examine mobility inequalities, access to essential services, and how socioeconomic factors influence travel behaviours.

High-frequency GPS data offers a powerful tool for interdisciplinary research. By combining GPS data with other contextual information, such as demographics or environmental conditions, researchers can develop comprehensive models that inform policy, improve public services, and enhance overall quality of life.

Combined GPS and PA/Health data

Integrating high-frequency GPS data with high-frequency physical activity and sedentary behaviour data, combined with Geographic Information Systems (GIS), provides a comprehensive approach to understanding how environments influence health. GIS allows researchers to spatially analyse exposure to health-promoting and health-harming environments, such as proximity to green spaces, air pollution levels, and access to walkable infrastructure. This combination enables detailed exploration of how environmental features shape movement patterns, physical activity levels, and sedentary behaviour, offering critical insights for urban planning and public health interventions.

In **urban planning**, GIS-enhanced GPS and activity data integration can pinpoint areas where environmental modifications—such as improving pedestrian pathways or increasing green space access—could encourage more active lifestyles. By overlaying spatial data with mobility patterns, planners can assess whether urban design elements support or hinder physical activity, helping to inform evidence-based policies to create healthier communities. GIS tools also facilitate the identification of transport and land-use patterns that influence sedentary behaviour and health outcomes.

From a **public health** and **epidemiological** perspective, combining GPS, physical activity data, and GIS enables the study of how different environmental exposures affect health behaviours and chronic disease risk. Researchers can analyse spatial-temporal trends to assess how individuals interact with their surroundings over time and determine associations between environmental factors—such as food availability, pollution exposure, or access to healthcare—and health behaviours. This approach helps to identify spatial inequalities and inform targeted health promotion initiatives.

A few specific research questions, across different fields, could include:

Public Health and Environmental Exposure

1. How do cumulative exposures to green spaces over time influence changes in physical activity levels in different population subgroups?
2. Does long-term exposure to air pollution impact the frequency and duration of outdoor physical activity (or other outcomes such as asthma)?
3. How do past health conditions (e.g., obesity, cardiovascular risk factors) predict current physical activity levels and mobility patterns?
4. Do past socioeconomic trajectories (e.g., income level, education) influence current access to and use of active environments?

Demographic and Socioeconomic Patterns

4. How do changes in employment status over previous sweeps relate to current sedentary time and travel patterns?
5. How do past household structures (e.g., living alone vs. with family) predict engagement with different types of environments?
6. How does the availability of, and exposure to, health-promoting environments influence long-term health disparities among different socioeconomic groups?

Sedentary Behaviour and Mobility Patterns

7. How do long-term commuting behaviours (e.g., active vs. passive modes) impact changes in sedentary time and activity patterns within cohort participants?
8. Are individuals with consistent exposure to mixed-use environments less likely to develop prolonged sedentary habits over time?
9. How do changing work-from-home trends influence sedentary behaviour patterns in urban vs. rural cohort participants?

Behavioural Change and Longitudinal Tracking

13. What environmental factors are associated with sustained increases in outdoor physical activity over multiple years in cohort participants?
14. How do life transitions (e.g., retirement, relocation) impact spatial mobility and physical activity behaviours over time?
15. Can longitudinal GPS and activity data predict future health risks, such as obesity or cardiovascular disease, based on spatial mobility and/or sedentary trends?

REFERENCES

Rollo S, Antsygina O, Tremblay MS. The whole day matters: Understanding 24-hour movement guideline adherence and relationships with health indicators across the lifespan. J Sport Health Sci. 2020 Dec;9(6):493-510. doi: 10.1016/j.jshs.2020.07.004. Epub 2020 Jul 22. PMID: 32711156; PMCID: PMC7749249.