Integrating wearable technology into population health monitoring systems for physical activity measurement: meeting report

25–27 November 2025
Utrecht, Kingdom of the Netherlands

World Health Organization

Integrating wearable technology into population health monitoring systems for physical activity measurement: meeting report

25–27 November 2025
Utrecht, Kingdom of the Netherlands

World Health Organization

Suggested citation. Integrating wearable technology into population health monitoring systems for physical activity measurement: meeting report, Utrecht, Kingdom of the Netherlands, 25-27 November 2025. Geneva: World Health Organization; 2026. https://doi.org/10.2471/B09745. Licence: CC BY-NC-SA 3.0 IGO.

Cataloguing-in-Publication (CIP) data. CIP data are available at https://iris.who.int/.

Design and layout by Inis Communication

Contents

Acknowledgements

This report presents outputs from the fourth in a series of meetings convened in support of the work of the World Health Organization (WHO) to develop new guidance on the use of wearable devices to measure physical activity and sedentary behaviours in national population health surveillance systems.

The meeting, held in Utrecht in November 2025, was jointly conducted by the Physical Activity team, Department of Health Determinants, Promotion and Prevention at WHO Headquarters and the Dutch National Institute for Public Health and the Environment (RIVM).

This meeting report was prepared by Dr Fiona Bull, Head of the Physical Activity Unit at WHO Headquarters, with technical support from Dr Mariken Leurs and Dr Juana Willumsen, of the Physical Activity Unit WHO Headquarters. Additional drafting input and review were provided by RIVM, Dr Annemarie Koster and Ms Inge de Wolf, and are gratefully acknowledged. Technical review by the session presenters was also highly appreciated, and all delegates were given the opportunity to review the report.

WHO extends its sincere thanks to all delegates who contributed to the meeting and to the development of this report. Appreciation is extended in particular to Dr Wanda Wendel-Vos and Dr Barbara Snoeker for their leadership in preparing for the meeting, and to RIVM for providing both financial and in-kind support to host and coordinate the meeting logistics, in close collaboration with Dr Mariken Leurs. Additional thanks are extended to the research teams who undertook rapid analyses to support the technical sessions, and to national government delegates for sharing their experiences and reflections. Collectively, the contributions of presenters and participants made the meeting both enjoyable and productive, and helped advance the global agenda on physical activity surveillance.

Executive summary

The World Health Organization (WHO) convened its fourth expert meeting on the use of wearable devices for measuring physical activity and sedentary behaviour in national population health surveillance systems on 25–27 November 2025 in Utrecht, Kingdom of the Netherlands.

The meeting brought together national government representatives, statistical agencies and scientific experts to review country experience with wearable devices in national surveillance, examine technical issues affecting comparability and interpretation, consider emerging evidence on wrist-worn measurement of cycling, and identify priority actions to support future integration of wearables into WHO's STEPwise approach to surveillance (STEPS) and other national monitoring systems.

Country experiences and key insights

Country experiences showed that integrating wearable devices into national surveillance systems is feasible but operationally complex. Approaches varied widely across countries, including differences in device type, wear location (wrist, thigh, hip), wear-time requirements, recruitment models, and data-processing methods. Most countries reported using wearable devices alongside questionnaires, recognizing their complementary roles. At the same time, countries differed in how wearable data were used within health monitoring systems, with some adopting device-based measures as primary indicators, while others used these data to generate behavioural insights to inform policy without replacing questionnaire-derived prevalence estimates.

This diversity in approaches reflected the need for future WHO global guidance and protocols to remain flexible, allowing adaptation to national contexts, resources and policy priorities. At the same time, it highlighted a central challenge: in the absence of clear guidance, wide methodological variation risks undermining data comparability over time, within countries and across countries.

Across almost all national experiences, representativeness was identified as the dominant challenge for wearable-based surveillance. Countries consistently reported lower participation among specific population groups, including older adults, people from lower socioeconomic groups, migrants, and individuals with poorer health or functional limitations. A range of strategies to improve participation and retention were described, including culturally tailored recruitment, community engagement, in-person contact, reminders, modest incentives and post-data collection participant feedback. However,

the evidence base on the effectiveness, equity and scalability of these approaches remains limited, and further empirical evaluation was recommended, including in low- and middle-income settings.

Cost and operational feasibility were also recurrent concerns. High device costs, device loss, postal delays, fieldwork complexity and the analytic burden associated with large raw datasets were frequently cited as financial and logistical barriers to wider adoption at national scale. Overall, the discussions reinforced that no single model can suit all country contexts, and that future WHO guidance should grounded in scientific rigour while enabling flexibility to accommodate national capacity, affordability, and implementation realities.

Technical considerations for wearable-based surveillance

Discussions of country experience highlighted the need to re-examine a small number of technical decisions that strongly influence wearable-derived estimates and their usefulness for policy.

First, **algorithm choice** emerged as the dominant source of variation across key metrics, including sedentary behaviour, step counts and moderate-to-vigorous physical activity (MVPA). Evidence showed that different analytical protocols applied to identical raw data can generate substantially different estimates, with direct implications for data comparability and policy interpretation. Participants therefore agreed that transparent, independently evaluated, open-source algorithms are essential.

Second, **wear location** was confirmed as a critical decision point. Estimates derived from wrist-, thigh- and hip-worn devices are not directly comparable, as each placement captures different movement signals and requires placement-specific analytical approaches. Evidence from machine-learning methods was shared, demonstrating progress in developing algorithms that may help harmonize data across different wear locations.

Third, **device brand** was found to matter less than previously assumed, provided devices are worn at the same location and raw data are analysed using standardized methods. Differences between commonly used research-grade accelerometers were generally small relative to variation introduced by algorithm choice or protocol decisions. This finding supports the importance of future global guidance remaining device-agnostic while focusing on defining minimum and optimal technical specifications.

Finally, **minimum wear-time criteria** were recognized as a key determinant of data quality and feasibility. Evidence reviewed suggested that the number of days required to produce stable population-level estimates may be fewer than previously assumed. However, uncertainties remain – particularly where protocols systematically restrict wear days – highlighting the need for further work to define feasible, defensible wear-time recommendations for national surveillance systems.

Measuring cycling

Countries with high cycling prevalence emphasized the importance of accurately capturing cycling behaviour within national surveillance systems, given its contribution to total physical activity and its relevance for transport, health and climate policy. Discussion focused on the extent of cycling misclassification when using wrist-worn devices and the implications for estimates of MVPA. The meeting reviewed recent advances in machine-learning approaches, including self-supervised models and signal harmonization methods, which show promising accuracy in free-living conditions. While these developments offer potential, delegates agreed that further refinement of the models is needed, as well as independent evaluation before wider use in national surveillance systems. Priority areas for further work include testing across more diverse populations and country contexts, inclusion of varied cycling modalities (including e-bikes), and assessment of implications for population-level estimates rather than algorithm accuracy alone.

Implications for WHO guidance and next steps

The meeting confirmed strong interest and momentum among governments, academic partners and WHO to advance the use of wearable devices for measuring physical activity and sedentary behaviour within national surveillance systems. Participants reaffirmed WHO's critical roles in convening, reviewing evidence and standard-setting to support countries, while recognizing that further methodological, operational and governance work is needed.

Priority actions identified focus on six interlinked areas. First, completion of WHO guidance on wearable devices for population surveillance was identified as a foundational normative deliverable, requiring further analytical work and consultation on: device specifications, benchmarking of algorithms, data governance and integration with self-report. Second, targeted research and development is needed to address key evidence gaps, particularly algorithm validation in diverse populations, including low- and middle-income countries and populations with atypical movement patterns. Third, the meeting recommended commencing a phased trial-and-adapt approach of demonstration studies as an intermediate step towards national adoption, ideally aligned with forthcoming STEPS surveys, supported by technical assistance, and – where needed – devices. Fourth, investment in developing national capacity – across fieldwork, data management, analysis and interpretation – was seen as essential for long-term implementation. Fifth, there is a need for strong communication resources to support policy-makers during periods of methodological transition and avoid undermining trust in established surveillance systems. Finally, alignment with evolving global physical activity guidelines was highlighted as critical to ensure future-proofing.

Overall, the meeting reaffirmed that wearable devices have significant potential to strengthen national physical activity surveillance and policy action, but realizing this potential will depend on coordinated global action, sustained investment, and clear guidance that balances scientific rigour with practical feasibility across diverse country contexts.

Background

Regular physical activity is essential for promoting both mental and physical health across all age groups. However, global levels of physical inactivity remain high: over 80% of adolescents and 31% of adults do not meet the levels of physical activity recommended by WHO.

To address this challenge, the Global Action Plan on Physical Activity 2018–2030 (GAPPA) *(1)* provides countries with a set of evidence-based policy recommendations (see Box 1). When implemented through a coordinated, multisectoral approach, these actions can help increase population levels of physical activity and contribute to achieving broader national health and development goals.

In 2018, Member States endorsed a global target to reduce physical inactivity by 15% among both adults and adolescents by 2030. Robust and regular population-level surveillance of physical activity is critical to track progress, assess implementation of GAPPA policy actions, and evaluate contributions to the prevention of noncommunicable diseases (NCDs) and other health, environmental, and societal outcomes. WHO technical support to countries includes providing guidance on instruments (such as surveys) and protocols to measure physical activity within national health monitoring systems, as well as assistance to advance policy implementation and monitoring.

At present, most countries rely on self-report survey instruments, which have known limitations such as recall inaccuracy, response bias, and high respondent burden (for example because of the complexity of questions, and length). Wearable digital devices – the market for which has seen a rapid increase in the past decade or so – offer a convenient alternative to self-report instruments. Developed for different wear locations on the body and with diverse functional capability, data processing methodologies, and prices,[1] these devices offer several advantages: they enable continuous, multi-day data collection across different movement types; are non-invasive; and place minimal burden on respondents.

Indeed, a small number of countries have now tested and/or incorporated a wearable device into their national monitoring systems in the quest for new ways to generate evidence that could potentially lead to new targets in future guidelines. However, there are technical, logistical, and financial barriers that currently limit the adoption of wearable devices and routine use at scale in national health surveillance systems.

For example, an unintended consequence of the expansion in the diversity in device wear location and functional capacity (reflected in the published research literature and across

[1] https://wearable-landscape.info/vision.

countries in both devices and use) makes it very difficult to assess the comparability of reported results – and the same is true for population estimates of physical activity based on wearable devices. And for the purpose of population health monitoring, there is currently no internationally agreed standard – nor WHO-endorsed guidance – on which wearable device(s) are appropriate, nor which data collection and analysis protocols should be used to report physical activity metrics.

To address these issues, WHO has convened a series of meetings to understand and scope the potential application of wearable technologies, the evidence base evaluating the performance of wearable devices, and related knowledge gaps. The first meeting (in November 2023) focused on current measurement methods and applications in adult populations *(2)*, while the second meeting (in May 2024) followed the same agenda but focused on adolescent populations *(3)*. Both meetings brought together scientific and technical specialists in measurement methods, wearable sensor technology, epidemiology, and population health surveillance of physical activity. A third meeting (in June 2025) built on the progress made and specifically set out to draft a set of wearable-device specifications for a defined set of priority metrics, and identify opportunities to address knowledge gaps and research priorities *(4)*.

Collectively, these meetings concluded that while it may be possible to make the transition to using wearable devices in monitoring systems for both adult and adolescent populations, and that the technology was getting relatively more affordable and feasible, several important logistical challenges and knowledge gaps remain. These issues are documented in the meeting reports and include: a lack of independent evaluation of analytical methods and machine-learning algorithms; limitations in training datasets (with poor diversity in the populations covered); and an absence of evidence from low- and middle-income country contexts. In addition, practical and logistical barriers to using wearable devices in national health monitoring are repeatedly raised, which include low response rates; sampling bias, leading to poor representativeness of the national population; and the costs and complexity of administering and handling very large and complex datasets.

These and other issues were discussed in further detail during the fourth physical activity measurement meeting, held in November 2025. The meeting comprised delegates with scientific expertise in the measurement of physical activity, population health surveillance, use of wearable device and machine learning methods for data analyses and representatives from national governments and statistical agencies with experience of using wearable devices in national health monitoring systems.

Objectives of the meeting

The meeting had three objectives:

1. To share and review country experiences on the use of wearable devices in population health monitoring systems.
2. To review the evidence evaluating use of wrist-worn wearable devices to measure cycling behaviours.
3. To discuss knowledge gaps and research priorities to support testing and scaling of wearable devices in population health monitoring systems.

This report summarizes the technical discussion held during the fourth physical activity measurement meeting. It serves to disseminate the discussion and recommendations arising from the meeting and to inform the development of technical guidance on the use of wearable devices to measure physical activity in national monitoring systems.

Box 1. Global policy calls for strengthening data systems and innovation

Strategic Objective 4 of the Global Action Plan on Physical Activity calls for all countries to strengthen data information systems to support policy- and decision making, tracking of progress and the promotion of physical activity by:

- Policy recommendation 4.2: "enhanc[ing] data systems and capabilities … to support regular population surveillance of physical activity and sedentary behaviour, across all ages and multiple domains; development and testing of digital technologies to strengthen surveillance systems; … to ensure accountability and inform policy and practice";
- Policy recommendation 4.3: "strengthen[ing] research and evaluation capacity to stimulate the application of digital technologies and innovation to accelerate the development and implementation of effective policy solutions".

The WHO Global Strategy on Digital Health 2020–2025 *(5)* also calls for advancing the use of digital technologies, including artificial intelligence (AI), to improve health including through health promotion, surveillance and behaviour change.

Section 1: Country presentations – learning from experience

A core objective of the meeting was to capture and discuss the lessons learned from the experience of eight countries[2] in using wearable devices to measure physical activity and other related behaviours (such as sedentary behaviour and sleep) in national population health, sport, or physical activity monitoring systems. A total of 10 approaches from these eight countries were presented at the meeting (two countries presented two approaches each, see Table 1). One additional country (the United Kingdom of Great Britain and Northern Ireland) shared a planned protocol for a feasibility trial of wearable devices within a national monitoring system scheduled for implementation in 2026 (not included in Table 1).

To facilitate the exchange of knowledge, support discussion and enable a comparison of all protocols and technical details used in each national experience, a summary table of their methods was completed using a standard template developed by WHO. Reports for all 11 examples are available on request.

Table 1 provides an overview of the key details from 10 completed national experiences, and further details on each are available in Annex 3. Overall, the protocols used in this set of country experiences show considerable diversity. Highlights of the comparative analysis are as follows:

- **Devices**:
  - Four approaches used an ActiGraph (various models) – an accelerometer worn either on wrist, hip or thigh.
  - Three approaches used thigh-worn devices – Axivity, SENSmotion and ActivPAL.
  - One approach designed and manufactured an accelerometer (UKK-RM42) worn on the waist (and wrist at night); these devices were borrowed by a second country to trial in their approach.
  - One approach used a pedometer (Yamax) worn on the hip but recently changed the protocol to allow the pedometer to be carried in a front pocket.
- **Frequency:**
  - Japan has completed annual assessment for 34 cycles over 36 years (1989–2024).

[2] Canada, Denmark, Finland, Japan, Netherlands (Kingdom of the), Norway, Sweden, the United States of America.

- Canada has experiences from eight (ongoing) 2-year cycles (beginning in 2007).
- Norway has experience from seven cycles (three among adults and four among children and adolescents).
- The Kingdom of the Netherlands had completed a pre-pilot in 2023 followed by a regional pilot study in 2024.
- All other countries had conducted between 1–3 cycles and most were conducted in recent years (that is since 2021).

- **Wear-time:**
  - The majority of approaches required a wear-time of 7 days; either continuously across 24 hours (in seven approaches) or worn during waking hours only (in one approach).
  - The Kingdom of the Netherlands requested a wear-time of 9 days (24 hours) and Japan requested only one day, only during waking-hours.

- **Data processing**:
  - Minimum wear-time for inclusion in analysis varied from one day to at least either three or four 4 days; some approaches employed additional criteria requiring a minimum number of hours of data collected per day or minimum step-count recorded per day.
  - Across the 10 approaches there were examples of the manufacturer's proprietary algorithm being used (in three approaches) and the development of "in-house" algorithms (in five approaches) of which most were open source.

- **Device administration**:
  - In all but one approach (Japan), devices were handed to participants during an in-person interview, conducted either at a centre or in the home (in four approaches) or posted to the individual's home after recruitment (in six approaches). In Japan the survey field worker handed the device and a survey to the participant, there was no interview.
  - All but one approach (Japan) used a reply-paid postal service for the return of the device; in Japan the pedometer was given to participants to keep, so no return was required.

- **Context**
  - In two approaches (Finland FinFit and Norway NNPAS), wearable device-based measurement of physical activity was implemented in the context of a national system that also assessed individual physical fitness.
  - In all other approaches the wearable device was used within context of a national health survey (i.e. health interview with questionnaire), and for many this also included collecting additional physical measurements (e.g. height/weight, blood pressure) and biochemical measurements (e.g. blood glucose).

The analysis reveals that there are multiple ways to use wearable devices as part of a national monitoring system, and so this wide variety of national protocols may limit the comparability of reported data and national estimates of physical activity. The current level of diversity between protocols, even across this relatively small number of countries, highlights the need for global guidance to support all Member States on the options and appropriate choice of device, methods, data analysis and reporting. The meeting revealed that global guidance to ensure adoption of the most effective approaches, strengthen data collection and use, and allow country comparability is now due. The general discussion acknowledged that while individual countries place focus most on monitoring their own population, the opportunity to make country comparisons is of added value. From a global perspective it was acknowledged that comparability of wearable device data across countries, and/or the ability to harmonize different datasets, would be essential for the future development of device-based global estimates of physical activity and tracking of changes over time.

Country presentations aimed to share reflections and learnings from practical experience and national contexts. Presenters were invited to address the following areas:

1. Briefly outline the context and monitoring system, project or survey, and outline government motivation to test or adopt a wearable device as part of national physical activity monitoring.

2. Identify the main challenges experienced in using a wearable device in your approach (for example issues related to protocols, infrastructure, organization, people, collaborations and processes).
3. Comment on response rates and the representativeness of the wearable-device dataset share any strategies employed to improve the response rates, compliance with protocols, and use of any statistical adjustments.
4. State the primary output metrics and comment on your experience of reporting data from both wearable-devices *and* self-report questionnaires, sharing any lessons on how best to communicate these data.
5. Provide key “take homes” and share any future improvement plans.

Table 1. Summary of selected characteristics across country approaches

Question[a]	1.1 System/study name	3.1 Device	3.4 Wear location	2.8 Wear time	6.4 Minimum wear time	6.5 Algorithm	1.7 No. of times
Canada	Canadian Health Measures Survey (CHMS)	Actigraph[b] wGT3x-BT	Waist/hip	7/24	≥ 4 days & ≥ 10 hours	Developed own in SAS (Accel) and in R (Actical)	8 cycles
Denmark 1	Danish National Health Survey (DNHS)	SENSmotion	Thigh	7/24	3 days & ≥ 20 hours +5 minutes walk/day	P[g] – Actimotus	1
Denmark 2	Moving Denmark 2020	Axivity AX3	Thigh	7/24 – RW[e]	≥ 4 days & ≥ 10 hours + 1WEND[d]	Developed own	1
Finland 1	Healthy Finland 2022–23	Actigraph GT9X Link	Wrist, ND[c]	7/24 – RW[e]	3 days – 24 hours	GGIR R-package, Hildebrand / van Hees algorithms	2
Finland 2	FINFIT STUDY	UKK – RM42	Waist, ND[c]+ Wrist at night	7/24 – RW[e] + switch[f]	≥ 4 days & ≥ 10 hours (+24h)	Own – UKK MAD-APE	3
Japan	National Health and Nutrition Survey (NHNS)	Yamax (pedometer)	Waist/hip (front pocket from 2024)	One day, waking hours	< 100 or ≥ 50 000 steps on the day were excluded	N/A	34
Norway	Norwegian National Physical Activity Surveillance (NNPAS)	Actigraph GTX3X+BT	Waist/hip	7/waking hours only – RW[e]	≥ 2 days & ≥ 10 hours	P[g] – Kinesoft	3
Sweden	Health on Equal Terms (HLV)	UKK – RM42	Waist, ND[c]+ wrist at night	7/24 -RW[e] + switch[f]	≥ 4 days & ≥ 10 hours + 1WEND[d]	Open source: MAD-APE	2
United States of America	National Health & Nutrition Examination Survey (NHANES)	Actigraph GT3x	Wrist – ND[c]	7/24	Various used inc. ≥1 day	Own – MIMS (open source)	2
Kingdom of the Netherlands	Accelerometer pilot	ActivPAL 3/4	Thigh	9 days/ 24 RW[e]	≥ 4 days /≥ 20 hours and ≥ 500 steps	P[g] PALtechnologies	pilot

[a] Refers to the corresponding question number in the template – available on request; [b] Actical device cycles 1–6, Actigraph cycle 7 in 2022; [c] ND: non-dominant wrist; [d] WEND: one weekend day; [e] RW – instructed to remove for water-based activity/shower/bath; [f] Switch: instruction for different wear location at night; [g] P: proprietary software

Key highlights from country presentations

Canada – Canadian Health Measures Survey (CHMS)

Context

- Evidence and experience from USA's National Health and Nutrition Examination Survey (NHANES) stimulated discussion in Canada and this, combined with published research showing poor correlation between self-reported and device-measured activity, raised government interest in the use of wearable devices.
- Large discrepancies in estimates have been found in both directions (i.e. self-report estimates higher than device-based estimates, and vice versa), and this – combined with long-standing concerns about social desirability bias and the difficulty of physical activity recall – has led to government efforts to implement a device with the primary goal of a more accurate, population-representative national surveillance and research system.

Device and protocols

- CHMS combines a household interview, physical health measures collected at mobile examination centres and accelerometry. It comprises a nationally representative sample of between 5000–6000 participants per 2-year cycle, with data collected from about 16 or 17 locations across the country; mobile examination centres are transported to collection locations and participants are selected from within 100 km of the mobile examination centre sites.
- Accelerometers are worn on the waist; Actical was used for cycles 1–6, with a change to ActiGraph for cycle 7 onwards; an additional change was made to the wear-time protocol, increasing from waking-hours only to 24-hour wear.
- Devices are handed out in person in the mobile examination centre and returned by pre-paid postal service.

Response rates and challenges

- There has been a declining participation in the survey over time, and especially in cycle 7 (post-COVID), but response rates appear to be improving again in cycle 8.
- A major challenge is the operational burden of conducting a system that involves the household contact, a mobile clinic visit and a wearable device protocol.
- The key point of attrition is people not attending the mobile clinic visit; non-compliance with accelerometer wear-time is less of an issue once participants agree to come to the mobile clinic visit.
- A second major challenge is dealing with non-respondent bias, and Canada uses a sophisticated weighting approach to make estimates nationally representative. Sample

survey weights are created for the full sample and the accelerometer subsample, and bootstrap weights are used to properly estimate the variance of the estimates.

Data processing and device comparability

- Changing the device from Actical to ActiGraph was a major challenge. In order to understand the impact Statistics Canada conducted a large cross-walk study (n=100); between devices, analyses showed good agreement and mean bias on moderate to vigorous physical activity (MVPA) estimates, but poor agreement and large absolute differences for light-intensity physical activity, sedentary time, and step counts. Therefore the conclusion was that direct comparisons and continuity of trend are not possible across the device changeover, except very cautiously for MVPA.
- Transparency about the uncertainty of estimates (variance) is very important and strict guidelines are in place to report standard error, coefficient of variation and confidence intervals.

Integration with self-report

- Canada continues to maintain its self-report questionnaire measure to provide information on physical activity by domains (transport, leisure, occupational) and uses the wearable device to provide quantitative data on total volume (guideline adherence) of sedentary time, light-, moderate-, vigorous-physical activity, and step counts.
- Wearable device and self-report measures are viewed as complimentary and both are seen as important in the context of the national surveillance system tracking 24-hour movement behaviours (physical activity, sedentary behaviour, sleep).
- Self-reported physical activity remains the core content on Canada's larger, population-based health surveys, which allow for better reporting on population subgroups (including equity-deserving groups, and to examine more local-level geographic distributions) that is necessary to inform policy and programmes.
- Device data (viewed as stable) have also been used to assess how self-reported data evolve over time in response to changes to the questionnaire instrument, and this has been very useful.

Future adjustments

- Canada is considering handing out the device at the household visit and collecting it during the clinic visit to reduce the burden of posting and to improve alignment (overlap) of the self-report and wearable-device-wear windows.
- Canada has shared Statistical Analysis System (SAS) code online for the analysis of Actical data, and code for use in the R statistical package is to be released in early 2026. These steps are important to standardize analysis, increase transparency of methods, and to support data users and researchers.

Canada – Key lessons

- National surveillance requires stability, rigorous weighting procedures to ensure national representativeness, and strong logistics.
- Changing the device used was extremely disruptive to trend data.
- Self-report data remains a vital piece of Canada's national physical activity surveillance programme. The crossover study was important to help understand the transition.
- Accelerometry adds major operational complexity; analyses are very complex and data users need analytical support, clear manuals and statistical packages (which Canada has produced and published).
- As interest in adopting a step-count as a surveillance target in Canada increases, the need to understand the impact of different bodywear location and different devices is vital. Whilst a step-count target offers a universal metric, it is viewed as problematic unless clarity on device and wear location is provided as part of future guidance (for example as benchmarks or standards).

Finland – Healthy Finland Survey (THL)

Context

- The Finnish Institute for Health and Welfare (THL) has been running health surveys since the 1970s.
- Wrist-worn accelerometry was integrated into the national health examination survey for the first time in the 2017 "FinHealth" survey, and again in the 2023 "Healthy Finland" survey.
- Healthy Finland collects a random sample of 60 000 participants for the questionnaire, 10 000 for health examination, and 2000 in the accelerometer subsample.

Device and data processing

- Healthy Finland uses the Actigraph GT9X-Link wrist-worn accelerometer, mailed to participants at home after obtaining consent during an in-person health examination; wear protocol is 7 days for 24 hours.
- Accelerometer data analysis is undertaken in collaboration with University of Jyväskylä.

Response rates and representativeness

- The health examination survey achieves a 59% response rate with the usable accelerometer data response rate achieving 55% (n=1086) in the accelerometer subsample; of these, 48% (n=953) returned the accelerometer including data and 43% (n=860) provided at least 3 days of useable 24-hour data.

- Non-response analyses are conducted for the whole sample, the health examination sample, and the accelerometer subsample data. To ensure population representativeness, population weighting is used to address non-response bias.

Integration with self-report

- A self-report questionnaire provided additional information on physical activity (i.e., domains and types of activity) and assessed sedentary behaviours. These data are used for evaluating adherence to physical activity recommendations, as they are from a larger sample with greater representativeness for reporting regional estimates.
- The data from wearable devices are used to provide information on total volume of physical activity, on light-, moderate- and vigorous-intensity activity, as well as sedentary time and sleep. These data are used for complementary analyses such as an assessment of 24-hour movement behaviours and "rest-activity" patterns.

Challenges

- Logistical administration is a challenge, specifically the alignment of timing the wearable device to start within a postal-delivery protocol; this often led to a loss of data as it was based on a typical 1–3 day postal delivery but it sometimes took longer. (In addition, a transport strike caused delays and cold weather reduced performance of device and battery life).
- Device problems include battery life, which is reduced when the device is unused for a long period of time.
- Device loss is another problem. Of the 150 devices provided, around 20 were lost, though some were recovered by postal tracking and phone contact with participant.

Looking ahead

- Next cycle planned for 2031 with in-person device administration at the clinic rather than using postal distribution.
- The purchase of new devices is required.
- There is a plan to extend analysis using the linked data from wearable devices with national registry datasets to examine associations of physical activity and sedentary behaviours with a wider range of health issues.

Finland – Key lessons

- Aligning the commencement of data collection from the device with postal delivery remains a challenge as delays can affect data collection by reducing battery life.
- A larger sample is needed for subgroup analyses.
- Handing out the device at the examination site would increase participation, reduce problems associated with delays, and improve data quality.

Denmark – Two experiences: National Health Survey and "Moving Denmark"

Context

- To date, two national Danish surveys – "Moving Denmark" in 2020 and the Danish National Health Survey (DNHS) in 2023 – have used accelerometry. Both studies should be regarded as pilot studies. Moving Denmark was a large-scale survey using device-based measurements in a subsample, and the DNHS 2023 was a mid-term surveillance study with a focus on health and health behaviour in general, using device-based measurements for only a subsample.
- As part of the ongoing DNHS, the National Institute of Public Health (University of Southern Denmark) applied for funding to conduct device-based physical activity measurement.
- Government interest in testing the use of accelerometers among the adult population was the result of a few factors: the positive experience from the Danish Health Behaviour in School Children (HBSC) project, which included device-based measures among adolescents; the recognized limitations of self-report data and the opportunity of devices to reduce recall and social desirability bias; the desire to test the reliability and validity of currently used questionnaires; and the interest in assessing the feasibility of integrating accelerometers in a large national-scale survey among adults.

Logistics and devices

- The SENSMotion device was used in DNHS 2023. This thigh-worn accelerometer with bluetooth-connectivity requires participants to download a phone app, pair their device, and upload their data to the cloud, preferably daily; also requires participants to complete a sleep diary and record time at school or work.
- Axivity AX3 (a thigh-worn device) was used in the Moving Denmark study.

Participation and representativeness

- 7000 participants were invited to take part in the 2023 DNHS, and around 1600 agreed. 3750 were invited to take part in the Moving Denmark study, and around 1500 agreed. There was selection bias present in both approaches, with those agreeing to participate being more educated, more likely to be employed and more active. One proposed solution was to oversample the harder-to-engage population groups, but this had resource and cost implications.
- Overall compliance with the protocol was quite good, probably because of the thigh-wear location and the device being waterproof so there was no need to remove it. In the DNHS, data were uploaded to the cloud continuously (participants had to activate Bluetooth).

Key lessons

- Clear instructions appeared to be essential. The need for a phone App adds complexity.
- Combining self-report (on physical activity context, etc.) and wearable device data is valuable and recommended as wearable device-based measures cannot stand alone.
- Postal logistics (e.g. delays, losses, damage) were a major issue; about 25% of devices were lost, creating a significant cost burden.
- Government funding was uncertain, with the high cost a major barrier.

Future potential adjustments

- The request to wear a device should be made at the start of the health survey, not at the end, because participants who have already invested time in answering questions are less likely to be willing to subsequently wear a device.
- A local pick-up/drop-off point for the device should be considered instead of "return by post" to reduce loss.
- Develop an automated system for sending reminders and simplify protocols.

Japan – National Health and Nutrition Survey

Context

- Japan's National Health and Nutrition Survey (NHNS) is legally mandated and has been conducted annually since 1948 under the Health Promotion Act.
- Japan incorporated interview-based monitoring of habitual exercise in 1986 and developed its first national physical activity guideline around this time.
- Strong noncommunicable disease policy momentum gathered after the 1964 Olympics and public interest in "steps" grew as part of the National 10 000 Steps Campaign.
- The establishment of a Japanese Industrial Standard (JIS) for pedometers in 1993 ensured the accuracy and reliability of the Yamax pedometer.
- Step count is a simple metric and suitable to communicate with policy-makers and the general public.
- Pedometers were introduced as part of the NHNS in the 1989 annual data collection for 34 cycles across 36 years (interruption in 2020 and 2021 due to COVID-19).

Protocol

- Data collection is carried out by 469 public health centres across the country, using multi-stage, stratified sampling of households; survey staff visit households in person and deliver devices and instructions.

- The Yamax pedometer is worn on the waist and participants are asked to reset the device in the morning, wear for 1 day (during waking hours), and self-record their step count in a diary. Survey staff later collect the log sheet. Participants are allowed to keep the pedometer as a gift so the device is not returned.
- In 2024 a change in the protocol requires the new Yamax pedometer model to be carried in participants' front pocket (of trousers, skirt or jacket); the new device stores up to 7 days' worth of data; however, national reporting will still use only one single day as recorded by the participant in the diary.

Strengths

- Introducing the pedometer was viewed as a relatively light additional burden compared to the already-introduced nutrition survey, blood sampling, and medical examinations.
- The protocol is very simple – 1-day measurement – requiring no device setup, no analysis, and no need for collection or return (participants keep the device).
- A pedometer was viewed as low cost, estimated at around ¥1000 each (about US$ 6), and Japan had set an industry standard for quality and performance.
- There was high acceptance by participants, with 78% of the lifestyle survey participants and 95% of the nutrition survey participants agreeing to wear the pedometer.
- Step-count is a simple metric for policy-makers, public communication and the setting of a policy target.

Limitations and challenges

- Maintaining representativeness is an ongoing challenge; the household response rate in 2023 was 48.6%.
- The pedometer provides only a step-count metric so it is not possible to report on MVPA or sedentary behaviours.
- The one-day measurement protocols may not reflect individual habitual activity.

Japan – Key lessons

- Simplicity enables long-term surveillance but limits depth of detail.
- Researchers have advocated for the adoption of accelerometers, but the government remains cautious about adopting accelerometer due to cost, logistics, and analytic burden; the current low-burden, low-cost system has achieved exceptional continuity and population acceptance.
- Moving to accelerometers would require major system redesign and resource investment.

Norway – National Physical Activity and Fitness Surveillance

Overview

- In 2001 a review of physical activity data in Norway found it to be of poor quality. In 2005, the government's national physical activity action plan (signed by eight ministers) mandated establishing national surveillance using "best possible methodology".
- The National Physical Activity and Fitness Surveillance system (NPAFS) has operated since 2005, surveying both adults and children/adolescents.
- The system uses a mixed design with repeated cross-sectional and some longitudinal components.
- The random sample is drawn from the national registry of adults aged 20–85 years. Total invited for NNPAS3 was n=14 569. Of these, 6089 (42%) agreed to participate, and of these, 5499 (38% of invited sample) provided a valid physical activity measure meeting the inclusion criteria for data analysis.

Device and protocols

- ActiGraph is waist-worn and mailed to participants and returned by post.
- The methodology used since 2005 has been consistent (although the actual ActiGraph models used have changed).

Response rates and representativeness

- The response rate is about 38% of those invited providing usable accelerometer data for analysis. Once participants agree, protocol adherence is high: around 14.5 hours/day; around 6.5 valid days (≥2 days minimum).
- Device loss is low (around 2–3%). Phone follow-up is used to retrieve unreturned devices.
- Persistent difficulty in recruiting immigrant groups (notably low uptake even after oversampling 10 000 people) and older adults.
- Strategies to improve participation include the use of gift cards and a lottery to win a bicycle. Providing direct feedback (summary of their data) to participants and use of text/SMS reminders significantly increased consent, device returns, and questionnaire completion.

Key implementation challenges

- The surveillance system was built "from scratch" and requires substantial resources.
- Requires a significant in-kind labour contribution from universities, for example 30% of a full-time position was required by the lead investigator to support the surveillance system.

- Need for complex data-management and analytic frameworks, which need to be continually updated.
- Coordinating multiple universities initially proved difficult but centralized collection has improved consistency.

Reporting and metrics

- Government primary reporting uses the device-based prevalence estimates (not self-report).
- Core metrics and outputs are: step counts, physical activity intensity distribution, adherence to physical activity guidelines and detailed time-trend analyses across 3 surveys in the adult population cycle.
- Norway continues to use a questionnaire (the International Physical Activity Questionnaire – IPAQ-short) but does not publicly report these data because of noncomparable prevalence estimates; communication of these differences is a challenge.
- Government is also interested in reporting on strength training and balance, which is creating methodological dilemmas.

Future considerations

- Funding remains uncertain, cycle to cycle.
- Norway is considering switching to a wrist-worn device but is concerned about trend comparability. It would require validation studies to bridge the change in device and wear location.
- New initiatives showing promise include recruitment through mosques and senior centres.
- The cost of devices (ActiGraph) remains high and future expansion of the system may need larger device stock.

Norway – Key lessons

- Long-term political commitment and financial support for device-based monitoring is essential for continuity.
- Representativeness remains the core challenge, not wear-time compliance.
- Consistency across waves is critical for time-trend analysis.
- Recruitment of minority and older populations requires targeted, culturally specific approaches.

Finland – FINFIT System

Overview

- The FinFit system began around 2009–2010 after the government and UKK Institute planned standardized measurement of physical activity, sedentary behaviour, and fitness.
- Four population-based surveys were run within each 4-year parliamentary term covering: school-age children (accelerometer + questionnaire); young adults; working-age adults (FinFit Working Age); older adults (FinFit Elderly). This is now combined into two 9-month data collection periods within each cycle. While the government is the client, the UKK Institute designs and executes the system.
- Finland places unusually high emphasis on physical function and physical fitness (cardiorespiratory and muscular); sedentary behaviour (including posture segmentation); and body composition (such as waist circumference and body mass index).
- This emphasis on fitness is the result of data that showed dramatic long-term decline in fitness in young men entering compulsory military service (i.e. a 400% drop in running performance over 40 years and a rise from 4% to 30% in the proportion of men with poor fitness). This is seen as having major economic and workforce consequences if trends continue.

Device and protocols

- Due to early challenges in finding a suitable device and algorithms, Finland developed its own accelerometer model UKK-RM42 – production cost is around €50 per unit.
- The device-wear protocol was evidence-based and involves wearing the device on the hip or waist during waking hours and switching to wrist-worn during sleep.
- New processing metrics (e.g. mean amplitude deviation) were developed.
- Measurement of physical fitness is undertaken at research centres.

Response rates and representativeness

- Strongly affected by COVID-19 in earlier cycles.
- Current cycle is improving with the use of local promotion, accessible data-collection venues, use of text message reminders, and providing detailed, personalized feedback.

Key outputs reported

- Steps, MVPA, light activity, sedentary behaviour (lying, reclining, sitting, standing).
- Fitness and body composition metrics.
- Cost analyses of low physical activity and high sedentary behaviour.

Finland – Key lessons

- A strong government interest in fitness and sedentary behaviour supports and shapes the national measurement strategy.
- Algorithm transparency and raw-data access are essential.
- Representativeness remains a core challenge despite high engagement and resources.
- Device cost, wear location, and analytic decisions heavily influence feasibility.

Sweden – National Survey: Health on Equal Terms

Overview

- Physical activity measurement has historically relied entirely on an annual/biannual national health survey (questionnaire) for people aged 20 years and above.
- This national survey uses a standard sampling of about 45 000 adults; some regions can increase the sample with financial support, which has led to sampling of up to 300 000 participants in some years.
- The survey uses a digital questionnaire with questions on lifestyle, health, physical activity and sedentary behaviour.

Device

- Sweden uses the UKK RM42 device (loaned to the Swedish government), introduced to the adult population subsample in 2018 (a wearable device had been introduced for a prior child population survey).
- Data collection is undertaken in collaboration with university partners.

Response rates and representativeness

- The response rate for the main survey has declined from around 58% to 39%; the wearable device subsample yield is far lower: only 25% agreed to wear device and around 16% provided valid usable data.
- Young adults are less likely to respond to the survey but more willing to wear devices – showing the opposite trend to that of older adults.

Metrics reported

- Sweden continued to report a prevalence of respondents meeting the 150-minutes per week guidelines using questionnaire-derived estimates for national statistics.
- Wearable-device data has been used to inform and guide the Swedish government's understanding of the "physical activity gap" (i.e. those not meeting the recommended levels), and in particular to show the proportion of the population with very high

levels of time spent in sedentary behaviours. This has been very helpful in "reframing" policy to focus purely on promoting moving more, and not only on achieving the MVPA recommendation.

Key challenges

- Sweden is still exploring whether to shift toward regular, wearable-based surveillance.
- Key questions include how to: increase response rates; achieve representativeness when low-SES groups are consistently underrepresented; determine how many days of wear are enough for usable prevalence; prevent devices being lost or worn by another person.

Sweden – Key lessons

- A subsample approach in a large national survey worked, but the yield of participants and data was low.
- Contrasting recruitment by age group requires tailored recruitment methods.
- Device data are powerful in engaging policy-makers – especially when they differ greatly to estimates from self-report questionnaires.

In addition to the presentations from countries with experience of implementation, two presentations shared learning from a recently completed pilot in the Kingdom of the Netherlands and a planned feasibility study in the United Kingdom of Great Britain and Northern Ireland.

Kingdom of the Netherlands – sensor-based physical activity measurement within the National Health Monitor (pilot study)

Presented by Dr Babara Snoeker

Context

- Long-standing concerns about the low accuracy of questionnaire-based physical activity measurement and its limitations for core indicators prompted renewed national interest in device-based measurement.
- The current Dutch system includes two surveys:

(1) a national survey – the National Health Survey (n≈9000) - implemented by Statistics Netherlands (CBS) ; and

(2) a municipal survey – the Municipal Health Services (GGD) Health Monitor (n≈500 000).

- Both surveys include the SQUASH – a questionnaire similar to the long form International Physical Activity Questionnaire (IPAQ-long) collecting data about physical activity in several domains (transport, household, work and recreation - including sports).

- Pilot studies conducted by the National Institute for Public Health and the Environment (RIVM), Statistics Netherlands, and Amsterdam UMC have provided insights into what helps or hinders participation in nationwide activity monitoring using accelerometers. The results show that collecting high-quality accelerometer data is possible with motivated staff and structured procedures, although clear role distribution, strong logistics, and digital support are necessary for successful large-scale implementation.

Device and protocols

- Statistics Netherlands selects a random sample from the Dutch population for the GGD Health Monitor.
- The GGD Health Monitor comprises a web-based survey and included a question to recruit participants into the wearable device sub-study; willing participants were required to confirm their preferred week for data collection which was used schedule the date for sending the wearable device to the participant by post.
- The ActivPAL device (thigh-worn) was used with valid wear-time criteria defined as ≥ 4 days of wear and ≥ 20 hours per day.

Response rates and representativeness challenges

- 125 495 participants from 2 GGD regions were invited to fill out the GGD Health Monitor in this pilot. The response on the questionnaire was 40 018. Of those, 6679 were willing to participate in the accelerometer pilot. Statistics Netherlands randomly selected 1100 to participate in the pilot. A total of 578 participants selected a week to wear the accelerometer, and of those, 506 had valid wear time of ≥ 4 days of wear and ≥ 20 hours per day (88%).
- A major operational challenge was the high loss of devices during fieldwork, attributed to the inexperience of the fieldwork company with handling accelerometers, in particular the limited effort given to contacting participants repeatedly to increase device return.
- Pilot results showed that men, younger adults (< 50 years), people on a low income, children of migrants, unmarried people, and people living in non-urban environments were less likely to complete the questionnaire. There was also a less willingness to wear the device among men, people on low income, older adults (> 70 years), people with a non-Dutch background, less-educated groups, those with worse self-reported health, those not meeting physical activity guidelines, and divorced individuals. These patterns created representativeness challenges in the pilot sample.
- Suggestions to mitigate the low recruitment rates among some groups includes over-recruitment of such groups, weighting of data based on census data, potentially linking data with cohort studies that have better capture of underrepresented groups (subject to alignment of physical activity metrics); targeted personal interviews with “hard-to-reach” populations; and continued methodological and logistical refinement to address selection bias.

Key challenges anticipated if the Kingdom of the Netherlands were to implement a wearable device

- Limited transparency of algorithms used to process accelerometer data (from the device manufacturer).
- Increasing (logistical) complexity with larger sample sizes.
- Managing the transition from questionnaire-only to mixed methods.
- Cost implications of device-based surveillance.
- Maintaining flexibility in indicators to allow adaptation when physical activity guidelines change.

Kingdom of the Netherlands – Key lessons and next steps

- The proposed plan for future use of wearable devices aims to integrate use into an existing national survey infrastructure, and not as a standalone data collection system. This integration will enable collection aligned with other health data and utilize existing infrastructure.
- The complexity of implementation was high, spanning logistics, algorithms, representativeness, costs and data governance – yet it is still considered feasible.
- Adopting methods to address the sampling bias is viewed as a critical issue and may require mitigation through complementary sampling strategies.
- Ensuring there is transparency in data processing and algorithms is essential, with an interest in using standard recommended approaches.
- Any future protocol is likely to propose a phased adoption whereby current national indicators reported using a questionnaire will move to a reporting of estimates from combined data sources (questionnaire and wearable devices), and in the longer term move towards reporting national estimates primarily derived from wearable devices supported by self-report measures (from a diary/questionnaire). This approach enables a balance of continuity while progressively improving measurement quality.

United Kingdom of Great Britain and Northern Ireland – Active Lives accelerometry feasibility study 2026 (planned feasibility study)

Presented by Drs Melvyn Hillsdon and Tessa Strain

Context

- Physical activity surveillance in England, as part of the United Kingdom of Great Britain and Northern Ireland, has evolved over several decades, from the Health Survey for England and the Active People Survey to the current Active Lives Survey for adults and children.

- Growing use of consumer wearables, combined with increased application in research, behavioural interventions and health care has increased government interest in moving to wearable device-based measurement.
- The United Kingdom Chief Medical Officers' Report (2021) recommended testing the feasibility of wearable devices in national surveillance.
- A window of opportunity opened during the renewal of 5-year contract for the Active Lives survey 2025–2030 and the testing of accelerometry was included.
- The aim of the feasibility study is to assess the opportunity to enhance understanding of physical activity patterns, not to replace existing indicators that are based on the self-report Active Lives survey. The collaboration between Sport England, the external survey implementor, and the academic partner is planned to ensure that all reported metrics are policy-relevant and interpretable.

Device and feasibility study protocols

- Active Lives uses a random household sampling from the postcode address file, to which it posts a written invitation to participate in and complete an online questionnaire (approx. 15–20 minutes); a paper questionnaire can be issued if requested.
- The Active Lives survey questions collect data on large number of different sports, including frequency, intensity and duration, plus a number of other attitudinal questions and health-related questions, as well as some environmental questions.
- Those who complete the online survey will receive the last question inviting participation in the accelerometry component; those who consent to be recontacted will be randomly selected and, if they fulfil the quota, be given a link; for these individuals contact details will be shared securely with the University of Exeter to implement the remaining protocol.
- The estimated sample size is 465, with the calculation based on answering the feasibility questions, not producing estimates of prevalence.
- GeniACTIVE, a wrist-worn accelerometer, will be posted to participants to be worn on the non-dominant wrist for 9 consecutive days; devices will be returned using pre-paid envelopes; also in that envelope will be a short questionnaire asking about their experience of wearing a device.
- Data will be processed by the University of Exeter, which will process the accelerometry data and return an anonymized data set back to Ipsos with the agreed set of metrics, and merged with questionnaire responses.

Primary objectives of the feasibility study are to establish:

- the proportion of individuals who consent to participate;
- the proportion that wears the device as requested;
- the proportion that returns devices with useable data;

- how participation and adherence vary by demographic subgroup, and how this affects data representativeness;
- cost per completed accelerometry data file returned.

Data processing and reporting – planned

- The study will use raw data rather than relying solely on behaviour-classification algorithms; proposed analysis and output metrics include measures of frequency, duration and intensity of active events; total volume of movement, fragmentation and breaks in movement; and bout distribution characteristics.
- Accelerometry data will be treated as complementary to self-report data collected through Active Lives, which will continue to capture non-occupational physical activity by domain, while accelerometers will provide objective data on total movement patterns.
- Integration of both data sources will be used to develop a new, digital physical activity evidence base and strengthen interpretation of self-report trends.
- Results on the cost per file will include the total cost including device procurement, distribution and return logistics, and data processing and management, and be used to inform future decisions on scalability and sustainability.

Intended outputs

- The feasibility study will determine whether accelerometry can be successfully integrated into an established national surveillance system with evidence on community acceptance of wearing a device for 9 consecutive days; operational feasibility at scale; bias and representativeness; and costs and sustainability.
- Experience and learning from a new model of collaboration whereby the implementation model puts the risk (and costs) for devices (including those lost) on the data collector partner, written into the collaborative agreement.
- Findings will inform whether and how accelerometry should be adopted into the future Active Lives surveillance system.

Summary of discussion

Plenary discussions enabled reflection across country experiences and highlighted common challenges, areas of convergence, and key conceptual considerations for the use of wearable devices in national physical activity surveillance. The main areas of discussion are summarized by thematic area below. Several issues – particularly representativeness, feasibility, standardization and communication – cut across multiple thematic areas and are therefore revisited throughout the discussion.

1. Response rates and representativeness of samples

One of the main topics discussed was the shared concern about bias in the data collected in study sub-samples using wearable devices. There was consensus that ensuring representativeness – not only increasing response rates – is essential for producing valid national prevalence estimates, and that this remains the most significant challenge across almost all national systems. Several countries noted that low participation contributes to persistent selection bias, even when large sample sizes are used. Countries consistently reported difficulties in engaging specific population groups, including older adults, individuals from lower socioeconomic groups, immigrants, and people living with chronic conditions or disabilities. There was also concern that those already active or interested in sport and exercise are more likely to participate, while those who are less active are less likely to do so, leading to overestimation of physical activity levels.

Methods to address this issue were explored. Oversampling was discussed as one potential solution but was recognized to increase both costs and fieldwork burden. Other approaches included developing culturally tailored recruitment strategies (as illustrated in Norway) and using community-wide promotion strategies (as used in Finland's FinFit). Both were reported to show promise, although no data were yet available to demonstrate effectiveness. Recruitment through in-person contact was generally associated with higher consent rates than postal or email approaches, and several countries reported likely future protocol changes in this direction.

It was repeatedly noted that comparisons of response rates between countries are difficult due to differences in calculation and reporting. It was suggested that a standardized format for reporting be developed to enable countries to share and compare response rates. The need for greater transparency in reporting study methods was frequently raised and recognized as a limiting factor for wider learning in this field.

Undertaking studies to test the effectiveness of different recruitment protocols was seen as useful, as was learning from recruitment and retention methods used in cohort studies. Delegates also discussed whether data from existing cohort studies could add value – for example, by increasing sample size or improving understanding of bias in national data. It was generally agreed that the potential use of cohort study data and methods should be explored further and potentially used in a complementary way alongside datasets collected through national surveillance systems. More experience is needed to understand how this can be implemented in practice. Such work is already commencing in the Kingdom of the Netherlands, led by RIVM and The Maastricht Study, and these experiences will be shared in forthcoming meetings.

2. Cost of devices and feasibility

Discussion then turned to financial and operational feasibility. While some countries benefit from stable government funding and political commitment, others face significant challenges in sustaining long-term investment. Partnerships and collaboration – including

substantial in-kind academic contributions – were seen as critical for supporting data collection, technical expertise and analysis.

A second significant area of discussion was the cost of devices, which was repeatedly cited as a major constraint affecting both sample size and government interest in adopting wearable-based measurement. Participants discussed trade-offs between device cost, sample size, wear-time protocols and delivery models (in-person versus postal). High device costs were viewed as a major barrier to scaling wearable-based surveillance, particularly in low- and middle-income countries. Device loss ranged from less than 2% to as high as 25%, affecting both budgets and representativeness. Japan's long-standing experience using a low-cost pedometer illustrated the value and appeal of simple, cost-efficient approaches.

The optimal frequency of wearable device–based surveillance was also raised. While more frequent measurement is desirable, many systems remain resource constrained, with reference made to WHO's indicative recommendation of at least one national data point every five years.

Delegates raised the potential use of data collected from personally owned consumer wearable devices (e.g. Fitbit or Apple Watch) as an alternative to deploying survey-specific devices. This prompted discussion of sampling bias, the influence of integrated feedback, data access and privacy, and proprietary data-processing methods. While acknowledging ongoing work in this area, further discussion was considered out of scope for this meeting.

3. How much standardization is needed?

Reflection on country experiences highlighted marked differences in methods, including device selection, wear location, data collection protocols and data processing. However, views differed on the degree of standardization required across countries or studies. Some participants strongly supported greater standardization, emphasizing that methodological differences affect comparability. Others suggested that differences could be addressed during data processing by developing and applying algorithms capable of harmonizing data collected using different protocols, including different wear locations.

Discussion focused on which elements should be standardized versus flexible. Four issues received particular attention: choice of wear location; device selection for a given wear location; wear-time protocols (including minimum days and valid wear-time requirements); and methods or algorithms for raw data analysis. Given divergent views and emerging evidence, participants agreed to dedicate a further session of the meeting to this topic, reported in Section 4.

Despite use of a WHO-developed reporting template, participants noted ongoing difficulties in comparing approaches and results across countries, with response rates cited as a clear example. Establishing a standardized approach to reporting study design and implementation details was again recommended, echoing calls made during the meeting on measuring physical activity in adults using wearable technologies (*4*). Inadequate

reporting of technical and methodological details in the scientific literature was also identified as a major barrier to transparency and comparability of national estimates.

Government representatives emphasized that their primary objective is to generate nationally representative data and monitor trends over time, with cross-country comparability being a secondary consideration. WHO noted that future global estimates of physical activity and sedentary behaviour using wearable-device datasets will require balancing these priorities, supported by agreed criteria for data inclusion and harmonization where appropriate.

4. Device administration, incentives, feedback and fieldwork logistics

Discussion then focused on practical implementation issues affecting feasibility and data quality. Although postal distribution of wearable devices may appear straightforward – particularly in high-income countries – experiences shared revealed the negative impact of postal delays, device losses and variable postal reliability. Some northern countries also reported reduced device performance due to cold temperatures affecting battery life. Hybrid administration models combining in-person distribution with postal return were common. While postal systems may be more scalable, in-person approaches were viewed as more effective for reducing device loss, improving participant instruction and controlling wear-start timing.

Participants noted that incorporating personal contact – either in person or by phone – can improve device return rates and shorten turnaround time, supporting more efficient device reuse and helping to maintain data collection schedules. WHO emphasized that in many low- and middle-income settings, postal distribution is often not feasible, and that in-person or clinic-based systems may be required.

Participants exchanged views on incentives to improve response rates. While some countries offer reimbursement or modest incentives, others face regulatory, ethical or sustainability constraints. Very limited experimental evidence was available to assess effectiveness, and no clear consensus emerged. Overall, incentives were viewed as potentially useful and best considered on a country-by-country basis.

There was strong agreement that participants should not receive feedback on activity levels during the wear period. However, delegates agreed that providing some form of post-data collection feedback represents good practice and may support engagement and goodwill. National experiences showed this to be feasible, although the scope and content of feedback varied. The limited evidence base on incentives and feedback strategies was identified as a knowledge gap requiring further evaluation.

5. Conceptual clarity on measurement objectives in national surveillance

This discussion focused on clarifying the purpose of wearable-device data within national surveillance systems. A recurring theme was the need to distinguish between population-level surveillance objectives and individual-level behavioural assessment. Participants

emphasized that national surveillance aims to track population averages and trends over time, rather than individual behaviour, and that this distinction should guide decisions on measurement priorities and standardization. There was strong agreement that surveillance outputs should inform and shape policy agendas. Some participants noted that future physical activity guidelines may evolve to reflect device-derived measures, reinforcing the need for an adaptive approach to wearable-based surveillance.

There was broad agreement that self-report questionnaires and wearable devices capture complementary dimensions of physical activity behaviour. Wearable devices provide more robust information on volume, intensity and accumulation patterns, while self-report remains essential for understanding context, purpose and domains of activity relevant to policy and intervention design. Participants acknowledged that systematic bias in self-report can also mislead decision-making, underscoring the need to support policy-makers in interpreting discrepancies between data sources.

Countries described differing conceptual models for the use of accelerometry, shaped by national surveillance systems, policy priorities and research capacity. The approach in the United Kingdom of Great Britain and Northern Ireland was presented as using wearable device-derived data primarily to generate additional insights that questionnaires cannot capture to inform policy – particularly on accumulation patterns, fragmentation of movement and social patterning. In this context, the wearable device was not being piloted to replace self-report prevalence estimates. In contrast, other countries (including Canada and Norway) reported using device-based estimates as their primary metric for assessing adherence to national physical activity guidelines. These examples illustrate that wearable devices may be adopted and used to serve distinct, though overlapping, surveillance purposes.

Device-based measurement was viewed as an important development for improving surveillance of physical activity and especially sedentary behaviour; it was also seen as likely to play an increasing role in sleep measurement as this area develops. Although most countries reported that governments value both data sources, some do not release results (or specific metrics) either from the self-report data or wearable-device data, largely due to challenges in communicating divergent estimates. Participants agreed on the need for clearer and more consistent communication of physical activity metrics for multiple audiences.

Related to, but distinct from, discussions on which wearable device to use in a national survey, the potential use of smartphones and consumer-device data was briefly raised. Although regarded as important to explore, WHO stated this dimension was out of scope for the meeting, given the significant technical, governance, privacy and equity challenges that currently limit feasibility at national scale.

6. Ethical and regulatory considerations

Ethical and regulatory issues were recognized as cross-cutting considerations. Participants noted that legal and regulatory frameworks – particularly the General Data Protection

Regulation (GDPR) in Europe – continue to constrain data sharing, secondary analysis and international collaboration. These constraints were seen as especially relevant for multi-country harmonization and reuse of wearable-device data.

There was broad agreement on the importance of robust safeguards to ensure ethical use, appropriate interpretation and protection of participant data, including transparency regarding data access, storage, and downstream use. Participants emphasized that ethical and regulatory considerations should be integrated into surveillance system design from the outset.

7. Communication, engagement and policy-facing interpretation

Across discussions, participants emphasized that successful use of wearable devices depends not only on technical feasibility but also on clear communication and engagement with policy audiences. Governments and decision-makers require accessible explanations of the purpose, value and implications of testing, adopting and integrating wearable-based measurement within existing surveillance systems.

There was strong agreement on the need for policy-facing products – such as concise rationales, policy briefs, technical notes and illustrative country examples – to support informed dialogue with governments. These materials were seen as particularly important for explaining how device-based measures complement, rather than replace, established self-report systems, and for avoiding unintended loss of confidence in long-standing surveillance approaches.

Participants also highlighted the importance of supporting countries to interpret and communicate results, including explaining uncertainty, methodological limitations, and discrepancies between device-based and questionnaire-derived estimates. Clear, consistent messaging tailored to different audiences was viewed as essential for appropriate use of wearable-derived data in policy and decision-making.

8. Role of WHO

Discussion highlighted WHO's critical role in providing leadership, coherence and technical stewardship in the evolving use of wearable devices for national physical activity surveillance. Participants strongly supported WHO remaining device-agnostic, while continuing to develop technical guidance that enables countries to adopt wearable-based approaches in ways that are scientifically robust, feasible and policy-relevant.

Priority areas for WHO leadership included defining minimum device performance requirements, promoting transparency in data processing and algorithms, and supporting the development of standardized analytic workflows and core protocol specifications. Participants emphasized that guidance should distinguish clearly between minimum and ideal standards, allowing national adaptation while maintaining coherence for international learning.

WHO was also seen as having an important convening and facilitating role, bringing together countries, researchers and regional offices to share experiences, build capacity and accelerate learning. This includes supporting harmonization where appropriate, facilitating access to technical expertise and third-party support, and assisting countries with implementation, data management and interpretation challenges.

Participants further noted the need for WHO guidance to explicitly address equity and inclusiveness, including the performance of devices and algorithms among populations with atypical or impaired movement, such as older adults and people living with disabilities. This was identified as a significant evidence gap and an area for future prioritization.

Overall, participants emphasized WHO's role in providing technical guidance and facilitating alignment and shared learning across countries, while recognizing the need for national adaptation.

Section summary

Across country experiences, participants highlighted shared challenges and emerging lessons in the use of wearable devices for national physical activity surveillance. Ensuring representative samples – rather than simply increasing response rates – was identified as the most significant and unresolved challenge, alongside concerns about cost, feasibility and operational complexity. Countries reported wide variation in devices, protocols and analytical approaches, with differing views on the degree of standardization required, reflecting diverse surveillance objectives and capacities. There was strong agreement that wearable and self-report measures serve complementary purposes and that clear communication is essential to support appropriate interpretation and policy use of results. Ethical, regulatory and data-governance constraints – particularly related to data sharing – remain important considerations. Participants emphasized WHO's role in providing device-agnostic technical guidance, facilitating shared learning and alignment across countries, and supporting flexible, evidence-informed adoption of wearable-based approaches that respect national contexts and priorities.

Section 2: Improving response rates and compliance: a key challenge to adoption

This section reports on the meeting's discussion on how to address the recurring issue of improving the representativeness of the sample recruited to a wearable device-based measurement, namely, how to optimize recruitment and ensure compliance with the wear protocol. Participants shared experiences from across their own national approaches.

Use multiple invitation methods to increase reach

A combination of a letter plus face-to-face and/or phone contact was seen as effective, especially for harder-to-reach groups, and can be adapted to the local context. In-person approaches (such as those used in the WHO STEPwise approach to surveillance, STEPS) allow direct introduction, explanation of purpose and demonstration of the device. In some contexts, digital demonstrations may be a suitable substitute. Opt-out (rather than opt-in) recruitment approaches may improve participation. Recruitment was reported to be easier when participants were already invested in a survey (e.g. after a household visit or attending a mobile exam centre). Overall, it was deemed important to minimize the number of steps or tasks required for the participant to respond and be recruited.

Provide clear and simple information about the study

To improve recruitment, ensure participants understand the purpose of the data and how it will be used by providing simple, non-technical explanations of what the device measures, and what it *does not* do (e.g. that is does not have any microphones, or track locations via GPS). Provide information in a format appropriate to the country situation and target populations (i.e. written information sheets, digital, and in all appropriate languages).

Correct framing of the study and purpose is vital and should emphasize the contribution to public health, a focus on daily activity (not "sport" or "exercise"), and that it is not aimed at monitoring individuals but the population as a whole. It was suggested to engage behavioural experts to help optimize the framing and develop effective messaging and communication materials.

It was also suggested it may be advantageous to ensure the interviewers/data field workers wear the device themselves, thereby modelling its use and potentially reassuring participants.

Provide clear guidance and demonstration of device use

When providing information on device fitting, its use and wear protocols, in-person explanations were considered the most effective, especially when the device was issued immediately after recruitment. This "immediacy" was considered to help participants understand the purpose and task required, and to reduce uncertainty and concerns. Use of instruction sheets and support materials such as links to short videos (for example, as used in the Kingdom of the Netherlands),[3] can help explain the use of the wearable device.

Staff visibly wearing the device themselves (for wrist-worn devices) and helping participants put on the device during an in-person hand-out can reduce later wear errors. Extra belts or straps can be helpful to allow for a clean replacement and/or when the strap gets wet.

Device comfort and convenience were seen as strong facilitators of better compliance. Wrist-worn devices were seen as desirable choice for this reason. However, it was noted that some sports require wrist worn devices to be removed. Waterproof devices and wear-time protocols that require 24-hour wear removes the risk of participants forgetting to put them back on in the morning or after water-based activities. Some suggestions noted that the appearance of the device may matter (if the device was visible), and therefore it may help if it "looked good" or could be discretely covered, although this may vary by cultural context.

Offer meaningful returns and incentives

In general, the use of incentives was seen as likely to increase recruitment and compliance. The incentive should be associated with completion of the study – e.g. correctly wearing and returning the device – and not linked in any way to the physical activity behaviours or whether the device recorded valid data, as malfunctions are not the participant's fault.

Incentives can include providing individual reports after the wearing period or broader health information (e.g. combined with other data from a health assessment). While it was generally considered that participants appreciate receiving feedback on their own data, it may also improve compliance with protocols. But the provision of feedback was viewed as a courtesy and a transparency tool, as well as potentially a motivational one.

Examples of effective incentives were proposed, including vouchers, lottery entries (e.g. the chance to win a bicycle), and small tokens for device return. All incentives must be legally and ethically approved and will vary across countries.

[3] Participant information video used in ActiveLIFE (https://youtu.be/OqYRWyvl58U).

Use reminders and offer access to ongoing support during the wear period

Providing a helpdesk, telephone hotline, or informational website for participants to use if there were problems or questions about battery charging (if this is needed), straps, skin irritation, or other device issues may be useful (especially if the device was delivered by post to the home address or by external fieldwork companies who may have less knowledge and training specific to physical activity and wearable devices).

SMS messages, phone calls or emails could be used to remind participants to keep wearing the device, and later to remind them when and how to return the device. It was noted that reminders did not work equally well everywhere. Also, although such efforts may not result in significantly better compliance, they do help maintain contact with participants which can be particularly useful to follow up on unreturned devices to reduce expensive device loss.

Provide interviewer training and pilot the process

Recruitment is enhanced through good interviewer training, both to ensure recruitment protocols are correct and consistent and that they are confident in handling common questions. This is especially important when external fieldwork companies are involved in data collection.

Piloting in "real-world" situations is essential to allow interviewers to practice and experience diversity of community questions and reactions to the device-recruitment protocols. Failure to invest in training materials and time for all data collectors can lead to low recruitment and device loss, as was the experience shared by several approaches.

Tailor recruitment strategies

Different recruitment strategies may be required to reach and engage some segments of the population and should be considered and pilot tested. Many settings find it harder to recruit older adults who tend to be less willing to wear accelerometers. Recruitment through specific settings or using information materials that show older adults wearing the device may be useful. Specific training on recruiting harder-to-engage groups should be developed and used to improve response rates. This may also apply to recruiting individuals who are less active, as they see themselves as not relevant. Careful framing, Q&A tip sheets, and images that do not suggest a high level of physical activity or sport performance may be helpful.

Section summary

The discussion highlighted that improving response rates and compliance is central to the successful adoption of wearable device-based measurement and remains one of the most challenging aspects of implementation. Experience across countries showed that multi-modal recruitment, combining written invitations with in-person or phone contact, is more effective than single approaches, particularly for harder-to-reach groups. Clear, simple and culturally appropriate communication about the purpose of the study and the function of the device is critical to building trust and participation, as is correct framing of the study as a population health activity rather than a focus on sport or individual monitoring.

Participants emphasized that immediacy, demonstration and interviewer confidence strongly influence compliance, with in-person explanation, visible modelling of device use by staff, and well-designed training materials reduce wear errors and device loss. Comfort, convenience and wearability – particularly for wrist-worn, waterproof devices with 24-hour wear protocols – were seen as important facilitators of adherence. Incentives and post-collection feedback were generally viewed as helpful for recruitment and compliance when ethically and legally appropriate, although evidence on their effectiveness remains limited. Finally, robust interviewer training, real-world piloting, and tailored recruitment strategies for under-represented groups were identified as essential investments to improve representativeness and protocol compliance before scaling wearable-based surveillance nationally.

Section 3: Technical device specifications and protocols: revisiting four key issues in wearable-based surveillance

This session revisited evidence on four methodological issues that emerged repeatedly in discussions of country experiences (see Section 2) and remained unresolved following the meeting on measuring physical activity in adults using wearable technologies held in June 2025 *(4)*. Further discussion was warranted due to the range of views expressed by experts and country representatives on the degree of standardization required across key technical features of wearable-based measurement. These included device selection and wear location, algorithm choice, and minimum wear-time criteria for analysis. Participants differed in their perspectives on what level of standardization is required to support comparability and which elements should remain flexible to allow adaptation to national context. These issues are central to the development of WHO guidance on how wearable devices are selected, deployed and analysed within national surveillance systems.

To support the discussion, selected delegates presented rapid overviews of relevant evidence and led focused discussions on how these methodological decisions influence the data collected and resulting prevalence estimates. It was noted that some decisions must be made prior to data collection (e.g. choice of device and wear location, wear-time protocol), while others relate to data processing and analysis and can be modified post hoc (e.g. algorithm selection, minimum wear-time thresholds and criteria for valid days). Although these issues are presented separately below, participants noted that certain elements – particularly wear location and algorithm choice – are closely interdependent, as algorithms are typically developed and calibrated for specific device placements. Across the four issues, evidence consistently indicated that not all technical decisions contribute equally to variation in surveillance estimates, with algorithm choice emerging as the dominant driver.

Issue #1: Does the choice of device matter for a given wear location?

This issue examines whether the brand or model of accelerometer meaningfully influences physical activity estimates when devices are worn at the same body location and data are processed using harmonized analytical methods. This question is directly relevant to

procurement decisions, scalability and long-term sustainability of national surveillance systems. Evidence presented consistently indicated that, *for a given wear location,* differences between commonly used research-grade[4] accelerometers are generally small when raw data are analysed using identical or harmonized processing pipelines. Across studies, variability attributable to device brand or model was substantially smaller than variability arising from other sources, particularly algorithm choice and data-processing decisions.

Key supporting evidence[5] included:

- a comparison *(6)* of three non-dominant wrist-worn research-grade accelerometers (ActiGraph GT9X Link, GENEActiv and Axivity) processed identically using GGIR. GENEActiv and Axivity demonstrated good to excellent equivalence across most outcomes, including average dynamic acceleration (< 10% difference). ActiGraph-derived acceleration was approximately 9–11% lower. Time spent in sedentary and light-intensity activity was equivalent across all devices, while MVPA agreement was high but not always within the ±10% equivalence zone.
- an evaluation of four non-dominant wrist-worn accelerometers *(7)* used simultaneously for 7 days and processed identically using GGIR. Estimates of sleep, sedentary behaviour, light activity and MVPA were equivalent across devices, with identical prevalence of meeting WHO physical activity guidelines. Agreement in behaviour classification exceeded 84% of daily time for all device pairings, and MVPA differences were small (1–8 minutes/day).
- a comparison *(8)* of two dominant wrist-worn devices (Axivity AX3 and Matrix 003) used by British and Chinese adults using a United Kingdom-trained machine-learning algorithm. Average acceleration showed strong equivalence (≤ 5%) across devices in both populations. Larger discrepancies observed for MVPA – particularly in the Chinese sample – were interpreted as reflecting limitations in algorithm generalizability rather than intrinsic sensor differences.
- reported findings *(9)* from the ProPASS consortium comparing three thigh-worn accelerometers (ActiGraph GT3X+, Axivity AX3 and activPAL Micro4). When analysed identically, absolute differences across devices were minimal across a range of behaviours, leading the authors to conclude that device-related differences were negligible when processing methods were standardized.

Where differences between devices were observed, participants noted that these were most often attributable to differences in applied algorithms, device placement nuances, or algorithm training and transferability across populations, rather than to sensor hardware itself. This interpretation was reinforced by meeting discussions and aligns with conclusions drawn in other published work that similarly attribute much of the apparent between-device variability to analytical rather than hardware-related factors *(10)*.

[4] Research-grade devices: specifically developed for research purposes and as such provide no feedback to participants and have enough battery life and storage capacity to collect, store and enable download of raw accelerations data (adapted from *(7)*).

[5] These additional study details were added after the meeting, and during the writing of this report.

Delegates agreed that evidence indicates that *for a given wear location*, the choice of research-grade accelerometer has limited influence on derived physical activity estimates when raw data are processed using harmonized methods. Differences historically attributed to device brand or model largely reflect analytical choices rather than sensor performance.

Summary of discussion

The policy-relevant implication is that WHO guidance can remain device-agnostic, advising countries to select any wearable device for a given wear location provided it meets minimum technical specifications (including access to raw triaxial acceleration data at an appropriate sampling frequency). Delegates agreed that comparability of prevalence estimates – within countries over time and between countries – depends far more on wear location, algorithm choice and the data-processing pipeline than on device brand or model.

Issue #2: Does wear location matter?

Although the WHO meeting on measuring physical activity in adults using wearable technologies held in June 2025 *(4)* reached agreement on recommending the non-dominant wrist as the preferred wear location for population surveillance, this issue was revisited during the present meeting in light of emerging country experience. Presentations in Session 2 demonstrated that hip- and thigh-worn accelerometers have also been implemented successfully in some national surveillance systems, prompting further consideration of the implications of wear-location choice.

Participants distinguished between two interrelated dimensions of this issue: (a) feasibility and participant burden, and (b) the impact of wear location on behavioural estimates and surveillance outputs. It was noted that these dimensions do not always point to the same "optimal" wear location, and that policy-relevant decisions require balancing both considerations.

a) Does wear location impact participant adherence and burden?

A primary rationale underpinning the earlier recommendation of the non-dominant wrist was evidence of greater participant acceptability and adherence compared with hip- or thigh-worn devices. Wrist-worn accelerometers are generally perceived as less intrusive and easier to wear comfortably across daily activities, and acceptable to both men and women – factors consistently associated with higher compliance, particularly in protocols requiring 24-hour wear. Several studies assessing user preferences have reported wrist-worn devices to be more desirable than alternative placements *(11, 12)*.

Beyond feasibility considerations, participants also noted a strategic advantage of wrist-worn placement in relation to the broader epidemiological evidence base. Many of the large cohort studies that are shaping contemporary understanding of physical activity, sedentary behaviour and health – including those likely to inform future guideline development – have adopted wrist-worn accelerometry. Alignment with this dominant evidence base

was viewed as advantageous for ensuring conceptual consistency between surveillance systems, research evidence and evolving policy guidance.

Overall, participants generally supported the wrist wear location as advantageous from a feasibility and participant-burden perspective, directly influencing response rates, completeness of wear-time data and the robustness of population-level estimates. At the same time, it was acknowledged that hip- and thigh-worn protocols have been successfully implemented in large cohort studies and in some national surveillance systems, as described in Section 2, and may remain appropriate in specific contexts or for particular measurement objectives.

b) Does wear location impact estimates of step counts, sedentary behaviour and MVPA?

The second and more technically complex question concerned whether different wear locations yield materially different estimates for key behavioural metrics, and whether any wear location can be considered "better" for population surveillance. Participants emphasized that the answer depends on the metric of interest. For example, thigh-worn devices may perform better for posture and sedentary behaviour, while hip- or wrist-worn devices may perform differently for activity intensity or step counts. The challenge for national surveillance systems is to identify a single recommended wear location that can reasonably estimate multiple metrics while remaining feasible at scale.

Evidence[6] on differences between wear locations

Multiple studies comparing wrist-, hip- and thigh-worn accelerometers worn concurrently demonstrate substantial differences in absolute estimates across wear locations in both laboratory and free-living conditions. The signals derived from different wear locations are not directly comparable by default, as each placement captures distinct movement signals that require placement-specific algorithms to produce estimates of the behavioural metrics of interest. Ideally, different wear locations and algorithms should be compared against a common reference standard, but the majority of studies include a single device/attachment site *(13)*.

Studies comparing across wear locations to a common to a reference standard

One study found that 24% of 21 wrist-worn algorithms estimated sedentary time within 10% (±0.94 hours) of the reference of a thigh-worn device, noting reasonable accuracy for sedentary time from wrist-worn devices and showing substantial heterogeneity in accuracy across algorithms *(14)*. In contrast, another study reported that non-dominant wrist-worn accelerometers substantially underestimated sedentary time compared with thigh-worn, posture-based measures *(15)*. For example, daily sedentary time was 63 minutes lower in the count cut-point analysis and 50 minutes lower by the Euclidean Norm Minus One (ENMO) cut-point. Limits of agreement were wide, indicating poor interchangeability of estimates between placements when using either the standard cut-point-based or ENMO-based

[6] Additional study details have been added during report writing.

methods. A third study found hip algorithms were overall more accurate compared to wrist-algorithms at estimating sedentary time compared to direct observation *(16)*.

And in a laboratory-based setting using direct observation as a criterion measure, a fourth study found thigh-worn devices to have superior sensitivity and specificity (> 99%) across intensity categories and breaks in sedentary behaviour, compared with hip- and wrist-worn placements *(17)*. Wrist-worn devices showed lower accuracy for MVPA classification and overestimated breaks in sedentary time.

Comparative studies without a reference measure

One comparative study without a reference measure compared hip- and wrist-worn ActiGraph GT3X+ devices in free-living older women *(18)*. It found moderate correlations between placements (Spearman's $r \approx 0.73$) but large divergences in absolute values. For example, the differences between wear location and various output metrics were around 6000 steps per day, around 30 minutes per day for MVPA, and up to 4 hours per day for sedentary time. Wrist-worn devices recorded nearly twice as many steps, approximately 30 additional minutes of MVPA per day, and up to four fewer hours of sedentary time compared with hip-worn devices.

Evidence from meta-analyses and comparative studies reinforces these patterns, with studies reporting that wrist-worn accelerometers record, on average, 3537 more steps per day and more MVPA and less sedentary time than waist-worn devices *(19)*. And others demonstrated systematic differences in outputs depending on wear location, particularly in free-living conditions *(20, 21)*.

Interpretation and limitations of the evidence

While these studies clearly demonstrate that wear location materially influences behavioural estimates, delegates noted that much of the existing literature focuses on correlations or relative agreement rather than the validity of absolute values using robust reference standards in free-living settings. Methodological heterogeneity across studies – particularly in algorithms, cut-points, reference measures and analytical pipelines – limits direct comparability and complicates interpretation.

Importantly, few studies explicitly assess the downstream implications of wear-location differences for national prevalence estimates or trend monitoring, which are the primary objectives of population surveillance. Several participants noted that consistency of wear location within a country over time may be more important for trend detection than strict harmonization across countries.

Emerging harmonization approaches

Given the scale of systematic differences between wear locations, participants discussed the potential of data harmonization approaches as a complementary strategy to strict

standardization. Emerging work from the ProPASS Consortium[7] provides proof of concept that raw data collected from different wear locations can be harmonized using advanced analytical methods.

One recent evaluation demonstrated that machine-learning approaches could produce statistically equivalent estimates of activity duration from wrist- and thigh-worn devices for most activity types when compared with ground truth, although some behaviours (e.g. stair climbing) remained challenging *(22)*. These findings suggest that emerging open source harmonization algorithms could help bridge differences between placements, though further validation across diverse populations and contexts is required.

Summary of discussion

In summary, the discussion of evidence confirmed that while wear location can substantially affect estimates of steps, sedentary behaviour and MVPA, these differences cannot be interpreted in isolation. Algorithm choice interacts strongly with wear location, and placement alone does not explain the discrepancies observed across studies. To date, the evidence does not support a single wear location that is optimal for all metrics of physical activity and sedentary behaviour. Current evidence suggests that thigh-worn devices provide more accurate estimates for posture-based outcomes (such as sitting, standing and walking), whereas wrist-worn devices offer clear advantages in feasibility, participant adherence and scalability for population surveillance. Participants agreed that further research is required, particularly comparative evaluation of algorithms against established reference standards and the development of harmonization methods that may help mitigate differences between wear locations. Given the current state of knowledge, WHO guidance on wearable devices for national surveillance systems should emphasize that the choice of wear location be guided by surveillance objectives, feasibility constraints and the need to maintain internal consistency over time.

Issue #3: Does the choice of algorithm matter?

Across all metrics of interest – sedentary behaviour, step counts and MVPA – the evidence reviewed at this meeting consistently indicated that algorithm choice and data-processing decisions are the largest source of variability in wearable-derived estimates, exceeding the effects of device brand (for a given wear location) and, in many cases, wear location itself. This conclusion emerged most clearly from studies applying different analytical pipelines to identical raw accelerometer data. The additional evidence[8] discussed on algorithm choice included the following.

[7] See https://www.propassconsortium.org/.

[8] Additional study details have been added during report writing.

For sedentary behaviour

One study compared sedentary-time estimates derived from wrist-worn (GENEActiv) and thigh-worn (activPAL) devices using 21 commonly applied algorithms, including both cut-point–based and machine-learning approaches *(14)*. Using the thigh-worn activPAL as the reference (mean sedentary time: 9.4 hours/day; 62% of the day), equivalence testing (±10%) showed that:

- 16 of 21 algorithms (76%) produced sedentary time estimates that differed more than ±10% from the reference.
- Nine algorithms (43%) showed substantial bias (≈1.85 hours/day or more).
- Only five algorithms met equivalence criteria, including two machine-learning approaches (Trost Extended, OxWearables) and three cut-point methods based on ENMO thresholds (GGIR ENMO 40 mg; Bakrania ENMO 32.6 mg; Fraysse ENMOa 62.5 mg).

These findings demonstrate that sedentary time estimates vary widely depending on analytical approach, even when derived from the same underlying sensor data.

For step counts

An evaluation applied five open-source step-count algorithms and one proprietary method (ActiLife), to three publicly available datasets with ground-truth step counts *(23)*. Despite using identical raw data, correlations between methods were low (as low as $r = 0.50$), reflecting substantial disagreement in estimated step volume. These discrepancies were partly attributable to differences in how a "step" is defined across algorithms. Machine-learning–based approaches generally performed better, achieving the highest F1 score (0.89 ± 0.11) and lowest mean absolute percentage error (8.6 ± 9%) across datasets, indicating their potential utility for population-level step-count estimation.

For moderate-to-vigorous physical activity (MVPA)

Large-scale analyses further illustrate the magnitude of algorithm-driven differences. For example, two studies derived very different estimates while analysing the same accelerometer data from approximately 90 000 UK Biobank participants *(24, 25)*. These analyses showed that defining moderate-intensity activity using a simple acceleration threshold (e.g. > 100 mg) yielded estimates of around 700 minutes per week of MVPA, whereas machine-learning approaches applied to the same cohort produced estimates closer to 175 minute per week. Such differences have major implications for estimating the prevalence of meeting physical activity guidelines.

Generalizability and population dependence

A further critical consideration is whether algorithm performance is population- and context-dependent. Recent multi-country analyses *(26)* have shown that machine-learning models performing well in one population (e.g. urban adults in the United Kingdom of Great Britain and Northern Ireland) may perform substantially worse when applied rural

adult populations in China. Together with existing evidence, these findings highlight limitations in the representativeness of current training datasets and the risks associated with applying algorithms beyond the populations in which they were developed.

While this issue has been most clearly demonstrated for machine-learning approaches, participants noted that similar limitations are likely to apply to cut-point–based methods, which are also derived from calibration studies conducted in a limited range of populations. There was strong agreement on the need for more diverse free-living accelerometry datasets with robust ground truth – particularly from low- and middle-income countries – to support testing, validation and continued algorithm refinement.

Summary of discussion

In summary, discussion of the evidence led to a clear policy-relevant conclusion: algorithm selection is the most influential technical decision affecting the comparability and interpretation of wearable-derived estimates. Differences in algorithm choice can result in substantially different estimates – most notably for sedentary time and MVPA – even when devices, wear location and raw data are identical. Participants also noted that transparency and standardized reporting of device specifications, wear location, algorithms and wear-time criteria are essential to ensure interpretability and secondary use of national surveillance data.

Participants therefore recommended that WHO guidance should address algorithm choice explicitly. There was clear agreement that WHO should not recommend or endorse a single algorithm, but instead focus on defining standards of performance, transparency and evaluation that any current or future algorithm must meet to be considered suitable for population-level surveillance.

Delegates suggested that guidance should prioritize:

- **clear performance standards** for key surveillance metrics, defined and maintained by WHO or another recognized, standard-setting entity;
- **independent evaluation against robust reference standards**, using agreed test datasets and evaluation frameworks;
- **transparency in data-processing pipelines**, including clear documentation of inputs, assumptions and outputs; and
- **open and equitable processes** through which algorithms – whether developed in academia, government or the private sector – can be submitted for independent testing and assessment against agreed criteria.

It was recognized that further work is required before such standards can be fully operationalized, including expanded evaluation in diverse populations, continued development of harmonization methods across wear locations, and the establishment of benchmark datasets for performance testing. These gaps reinforce the importance of

an iterative "test and adapt" approach to standardization, led by evidence and supported through WHO's convening role and technical stewardship, rather than through prescriptive endorsement of specific analytical methods.

Issue #4: Do wear-time criteria matter?

Minimum wear-time requirements determine which participants are included in analyses and therefore directly influence sample size, representativeness, feasibility and cost. At present, there is no consensus on several key aspects of wear-time criteria for wearable device–based surveillance, including:

- the minimum number of days required for computing population-level estimates;
- whether wear days must be consecutive;
- whether inclusion of a weekend (non-work) day is required; and
- how to handle partial days or irregular wear time.

A central point raised in discussion was that the appropriate wear-time requirement depends on the purpose of the analysis. In particular, the number of days required to produce *stable population-level means or prevalence estimates* may be substantially fewer than the number of days required to achieve high *within-person reliability*, which is often assessed using intraclass correlation coefficients (ICC). Reliability thresholds derived from ICC address individual-level ranking, whereas surveillance systems are primarily concerned with unbiased estimation of group means and prevalence.

Evidence[9] on wear time for population-level estimates

To inform this distinction, evidence from large-scale secondary analyses of NHANES 2003–2006 was recalled *(27)* which directly addressed the question of how many monitoring days are needed to produce *stable group-level estimates of a 7-day mean*, using data in youth (n=986) and adults (n=2 532) with 7 valid wear days. Using bootstrapping and simulated missing-data protocols, the study compared a wide range of alternative monitoring schemes (random days, forced inclusion of weekend days, and combinations thereof) against the full 7-day benchmark. Key findings included the following:

- Stable group-level estimates of MVPA, light activity and total activity could be obtained from as few as one randomly selected monitoring day, with mean bias close to zero across 1–6 randomly selected days.
- Non-random inclusion of weekend days introduced systematic bias in group-level estimates, with both over- and under-estimation depending on the protocol used.
- The study demonstrated that conclusions based on ICC-derived reliability (focused on individual consistency) differ fundamentally from conclusions based on stability of group means, and that ICC-based approaches are not well suited to informing population surveillance protocols.

[9] Additional study details have been added during report writing.

These findings reinforce that, for surveillance purposes, randomness in the selection of wear days is more important than duration per se, and that requiring specific days (e.g. weekends) may inadvertently bias prevalence estimates. However, delegates noted an important limitation in directly applying these results to structured surveillance protocols such as STEPS. While one study *(27)* demonstrated that randomly selected monitoring days can yield unbiased group-level estimates, it does not address whether protocols that systematically restrict wear time to the same fixed days for all participants (for example, only days 1–2 of a survey cycle) would produce similarly unbiased estimates. In such designs, bias may arise if participants interviewed later in the fieldwork cycle differ systematically from those interviewed earlier, or if activity patterns vary by day of the week in non-random ways. This distinction was viewed as particularly relevant for STEPS-style fieldwork, where operational constraints may limit the ability to randomize wear days across participants.

Evidence based on ICC-focused approaches

Delegates also discussed evidence using ICC-based methods, including recent analysis from UK Biobank *(28)* which used missing-data simulations to assess how many hours or days of wear were required to maintain acceptable reliability (e.g. ICC ≥ 0.90) for summary metrics such as vector magnitude. These analyses suggested that approximately 72 hours (≈3 days) of wear may provide acceptable reliability in high-income adult populations. While such findings are informative, participants emphasized that ICC-based reliability addresses a different question – namely the consistency of individual measurements across days – and should not be interpreted as a direct requirement for estimating population-level prevalence. The applicability of these findings depends critically on whether wear days are randomly distributed across participants rather than systematically constrained by fieldwork design.

Summary of discussion

In summary, delegates agreed that minimum wear-time criteria remain an unresolved but critical design choice with major operational implications. The discussion highlighted the following:

- Fewer wear days may be sufficient – and defensible – for estimating population-level means and prevalence, provided days are sampled randomly.
- Wear-time requirements derived from ICC-based reliability studies should not be applied uncritically to surveillance contexts.
- Further empirical work, including leveraging secondary analyses of existing datasets, is needed to assess minimum wear-time requirements for specific metrics (MVPA, steps, sedentary behaviour), across diverse populations and settings.

Participants agreed a rapid review of existing evidence and conducting analyses of large datasets to test alternative wear-time scenarios (e.g. 1, 2, or 3-day protocols) under realistic surveillance constraints be undertaken. Given the substantial cost and feasibility

implications, this was viewed as a priority area where evidence-informed guidance could have immediate practical value for countries planning wearable-based surveillance.

Section summary

Across the four technical issues reviewed – device choice, wear location, algorithm selection and minimum wear-time criteria – the discussion consistently highlighted that decisions which appear "technical" have substantial implications for feasibility, comparability and interpretation of national physical activity surveillance data. While countries expressed interest in standardization to support comparability, there was equally strong recognition of the need for flexibility to accommodate differing surveillance objectives, fieldwork constraints and national contexts. The evidence reviewed shows that not all sources of variation are equal, and that prioritization is essential in developing practical and credible guidance.

Overall, the discussion confirmed that algorithm choice is the dominant driver of variation in wearable-derived estimates, exceeding the influence of device brand and, in many cases, wear location itself *when raw data are analysed using appropriate methods*. For a given wear location, differences between commonly used research-grade devices are small when harmonized analytic pipelines are applied, supporting a device-agnostic approach to guidance.

Wear location, however, remains a critical structural decision. Estimates derived from different placements (e.g. wrist, thigh, hip) are not directly comparable by default, as each captures different movement signals and requires placement-specific algorithms (and, in practice, should be treated as measuring different constructs unless harmonization methods are explicitly applied). Thigh-worn devices provide greater validity for posture-based outcomes, while wrist-worn devices offer clear advantages in feasibility, adherence and scalability for population surveillance. Bridging these placements requires access to raw data and validated algorithms or harmonization methods; without these, differences in wear location will translate into systematic differences in estimates. No single placement is optimal for all metrics, reinforcing the importance of aligning wear-location decisions with surveillance objectives and maintaining internal consistency over time.

The discussion on wear-time criteria further emphasized that the purpose of surveillance matters. Evidence suggests that the number of days required to achieve stable population-level means and prevalence estimates may be substantially fewer than those derived from individual-level reliability analyses based on intraclass correlation coefficients. However, important uncertainties remain, particularly regarding structured protocols that systematically restrict wear days under real-world fieldwork constraints. Together, these findings underscore that WHO guidance should avoid prescriptive technical rules and instead focus on clarifying principles, performance standards and decision frameworks – supporting countries to make transparent, evidence-informed choices while enabling comparability, learning, and adaptation as methods and evidence continue to evolve.

Section 4: Measurement of cycling behaviours

A second core objective of this meeting was to review evidence on the performance of wearable device-based measurement of cycling. Specifically, the session aimed to review the evidence on how well wrist-worn devices perform compared with a ground truth measure and alternative wear locations such as the thigh. The measurement of cycling is an important issue for those countries where cycling is a popular form of physical activity for transport, exercise or recreation. This is particular true in the Kingdom of the Netherlands, where national data show that over 25% of all trips are made by bicycle *(29)* – higher than in any other country in the world *(30)*. Self-reported data indicate that adults (aged 18–65) cycle on average more than two hours per week, including both commuting and leisure-time cycling *(31)*. Given these national data, the Dutch government is particularly interested in how well wearable devices will detect cycling behaviours so that these minutes of at least moderate-intensity activity are accurately included in outcome metrics reporting prevalence of meeting the adult physical activity guidelines.

This session was supported by four presentations and aimed to:

1. assess the extent and strength of evidence evaluating wrist-worn wearable device-based measurement of cycling behaviours;
2. review the measurement performance of open-source machine learning algorithms of cycling behaviours;
3. identify knowledge gaps and create a list of research needs;
4. scope opportunities to conduct research to address knowledge gaps through collaborations, and the use of existing or new datasets.

Presentation 1: Understanding the public evidence on cycling-mode detection using wrist- and thigh-wearable devices

Presented by Dr Daniel Fuller

Overview

This presentation provided a rapid ("flash") review of the evidence on how accurately cycling can be detected from wearable devices – particularly wrist-worn accelerometers – using machine-learning models. Transport-mode detection (e.g. walking, cycling, standing, riding

public transport) is a rapidly advancing field with potential value for national physical-activity and mobility surveillance.

Methods

A flash review[10] was conducted to identify published studies meeting the following criteria: used a wrist-worn accelerometer; included cycling as a classification category (stationary and/or mobile); applied at least one machine-learning model; and reported a clear performance metric.

From 15 initially identified articles, six met inclusion criteria. Key data extracted included participant characteristics, device specifications, data-collection frequency, model types, classification categories, and performance metrics. Some studies also compared wrist-based classification with thigh- or hip-worn sensors.

Key findings

- Sample sizes varied widely (10–600 participants) and showed considerable heterogeneity in study design and populations.
- All included studies used Random Forest models,[11] with many exploring additional traditional machine-learning approaches.
- Classification tasks typically involved 4–10 activity categories, most commonly walking, sitting, standing and cycling.
- The average accuracy for detecting cycling from wrist sensors was 72% (range 46%–95%).
- Evidence comparing wrist to hip/thigh devices was inconclusive (two studies favoured thigh/hip; one favoured wrist).
- Reporting of key modelling steps – such as hyperparameter tuning, cross-validation procedures and dataset availability – was generally limited.

Key messages

The current published evidence suggests that wrist-based cycling detection is feasible but variable, with accuracy influenced by study design, model choice, device placement, and diversity of activity categories. Importantly, the reviewed studies largely relied on older machine-learning techniques. The presenter noted that newer approaches – particularly self-supervised learning and deep learning – are likely to provide substantial performance improvements and could help standardize transport-mode detection for public-health surveillance applications. Overall, published evidence provided some promising indications but was viewed as insufficient to inform policy decisions.

[10] For more details on novel "Flash" review methods see https://teaminteract.ca/flashreviews/.

[11] A Random Forest is a machine learning model that works by using many decision trees, collectively called a "forest," to make a prediction. Decision trees can be visualized as a flow chart of conditions that can be used to most accurately classify or predict outcomes.

Presentation 2: High-performance detection of cycling from free-living wrist accelerometer data using a self-supervised algorithm

Presented by Dr Laura Brocklebank, on behalf of co-authors Drs Aidan Acquah, Ben Maylor, Vic O'Callaghan, Aiden Doherty

Overview

This presentation introduced ActiNet, a new open-source, self-supervised machine learning algorithm designed to detect cycling from free-living wrist-worn accelerometer data. This new work was undertaken to respond directly to the need for more accurate, scalable transport-mode and activity-type detection, especially for integration into population-level physical activity surveillance systems.

Methods

- ActiNet was trained and evaluated using the CAPTURE-24 dataset, in which 151 adults (aged 18–91 years) wore an Axivity AX3 wrist accelerometer for 24 hours along with a wearable camera providing ground-truth labels for activity types, including cycling.
- Model performance was evaluated using five-fold cross-validation, with accuracy used as the primary metric to assess agreement between the ground-truth ("true") and model-assigned ("predicted") labels.
- In addition, an external validation currently underway using the English Longitudinal Study of Ageing (ELSA) provided a dataset of 82 adults ≥ 50 years using an identical sensing and wearable-camera protocol.
- Both labelled datasets are publicly available, enabling reproducibility and further method development.

Key findings

- The CAPTURE-24 dataset included 3114 30-second windows labelled as cycling.
- Preliminary results indicate that 95% of cycling was correctly identified by the ActiNet algorithm, representing a substantial improvement over an earlier random forest model which achieved 88.7% accuracy.
- ActiNet worked well for detecting outdoor cycling with a wrist-worn accelerometer, but not for detecting indoor or stationary cycling.
- Data show that 87% of cycling was MVPA and 13% was light-intensity physical activity.
- Upcoming additional analyses will assess per-participant accuracy, sensitivity and specificity, and performance on both internal (CAPTURE-24) and external (ELSA) datasets to assess generalizability.

Key messages

These preliminary findings indicate that the ActiNet algorithm showed good performance for detecting cycling in free-living conditions. The study demonstrates that self-supervised machine learning substantially enhances cycling detection accuracy from wrist-worn accelerometers – an important advancement for measurement of moderate-to-vigorous physical activity in free-living conditions. The open-source nature of ActiNet, combined with publicly available training datasets, provides a strong foundation for transparent, scalable and globally adaptable algorithms, with clear relevance to future WHO guidance on wearable-device based surveillance.

Presentation 3: Leveraging thigh-movement patterns to improve wrist-worn cycling detection using GAN-based signal harmonization

Presented by Dr Matthew Ahmadi on behalf of Drs Hossein Moeinzadeh, Alcion Popp, Manos Stamatakis

Overview

This presentation described an advanced machine-learning approach that uses Generative Adversarial Networks (GANs)[12] to translate thigh-based movement signatures into the wrist-signal domain. The study aimed to overcome the challenge that wrist and thigh sensors produce non-comparable signals, particularly for cycling, where in the thigh provides a strong biomechanical signal while the wrist does not. Improving wrist-based cycling detection would help achieve better interoperability across studies using different device placements.

Methods

- Data were collected from 75 adults in Australia (mean age 56 years; 54% female), each wearing both wrist and thigh Axivity AX3 devices and video-verified ground truth (with an average of approximately 6 hours of data per participant); 7 cyclists and 8.4 hours in total of cycling recorded.
- The research used a two-stage pipeline:
    1. Upstream signal translation using a GAN, trained to generate a harmonized wrist signal that captures informative patterns typically seen at the thigh.
    2. Downstream activity classification using a stacked Random Forest and Long Short-Term Memory (LSTM) model.
- Models were trained using 80/20 cross-validation.

[12] GANs help machines to create new, realistic data by learning from existing examples.

- External testing used the 151 participants from the CAPTURE-24 dataset, also with paired wearable and camera data; 42 cyclists with 24.4 hours of cycling recorded.
- Performance was evaluated via sensitivity, precision, F1-score, confusion matrices, Bland–Altman plots, and equivalence testing comparing estimated versus ground-truth cycling duration.

Key findings

- Internal (using Australia data) cross-validation: preliminary results for the wrist device showed a sensitivity of 92.5%, precision of 97.3%, and an F1 score of 94.8%.
- External (using CAPTURE-24) testing: sensitivity: 87.3%, precision: 99.2%, F1-score: 92.9% and confusion matrices showed robust discrimination of cycling from other behaviours across diverse activity contexts (86.4%).
- Duration-agreement analyses: Bland–Altman results showed minimal mean bias (approximately minus 1.6 minutes with 5% CI 7–10 mins) and acceptable limits of agreement.
- Equivalence testing indicated results that were *not statistically different* from ground truth, but not fully equivalent within the predefined equivalence zone (± 10%), suggesting room for optimization.
- Most misclassifications occurred between cycling and walking, likely due to similarities in movement patterns.

Key messages

This new work demonstrates GANs offer potential methodology for analysis of data from wrist-worn devices to detect cycling with comparable strength compared with thigh-based movement signals. Employing an approach that uses learning from thigh-derived movement patterns and mapping onto wrist-derived signals can substantially improve the accuracy of wrist-based cycling detection while supporting greater interoperability across cohorts and study protocols. Future improvements are expected through expanding training datasets (e.g. more cyclists, longer cycling observations), refining harmonization models, and exploring applicability to a broader set of movement behaviours.

Presentation 4: Independent performance evaluation of an open-source wrist algorithm for cycling detection

Presented by Dr Sarah Kozey Keadle on behalf of the Wrist Algorithms: Verification, Evaluation and Surveillance (WAVES) Research Initiative

Overview

This presentation reported an *independent* evaluation of the open-source ActiNet self-supervised algorithm for detecting cycling from free-living wrist-worn accelerometer data. The work forms part of the WAVES (Wrist Algorithms: Verification, Evaluation and Surveillance) initiative and applies the PhasED Framework for algorithm evaluation (Phase III: performance evaluation in naturalistic conditions). Evaluation was conducted comparing wrist-worn ActiNet to a primary reference measure of video-recorded direct observation. As a comparator to provide context for accuracy, thigh-worn activPAL data were also compared to direct observation.

Methods

- Data were drawn from two studies (ACT24 with 24 subjects, and AM with 25 subjects), each including healthy adults aged 18–80 years wearing both a wrist device for 7 days (GENEActiv or ActiGraph GT3X+ at 80 Hz), and a thigh-worn activPAL (CREA v1.3) device over the same period.
- Participants also completed two video-recorded observation sessions (2–3 hours each).
- Trained research assistants coded posture, movement type and intensity, including cycling, using a refined direct observation protocol.
- To compare metrics directly, the team converted direct observation labels to 1-second epochs, converted ActiNet's 30-second windows into 1-second estimates, aligned both with activPAL output, and calculated accuracy measures (precision, recall, F1,[13] balanced accuracy) and inspected false-positive/false-negative patterns.

Preliminary results

- Used a total of 178 hours of labelled data but only 2.5 hours of cycling behaviour recorded – a level lower than used in the CAPTURE-24 study but the richness of these data may help offset the low number of hours cycling.
- Within the direct observation verified subset, four participants engaged in four cycling events across a range of modalities: mountain-bike trail riding, cargo-bike commuting, road-cycling exercise, and stationary indoor cycling.
- Overall, ActiNet was found to be accurate (99.5% balanced accuracy) compared to direct observation, with an F1 of 0.85. For example, results showed for stationary indoor bike:

[13] F1 score is a machine learning evaluation metric that measures a model's accuracy.

direct observation (min) 11.1 and ActiNet (min) 11.0 and for road bike / exercise: direct observation DO (mins) 91 and ActiNet (mins) 96.

- ActiNet detected all cycling events (recall = 0.99), with some false positives (precision=0.74), all of which occurred during other MVPA such as walking while pushing an object like grocery cart or stroller – a known challenge in wrist-based classification.
- ActivPAL, was also accurate overall (91.2%) and F1 of 0.89. By contrast, activPAL showed almost no false positives (precision=0.98), but missed some cycling periods (false negatives as reflected in a recall of 0.85) – these were moments involving coasting or limited pedal motion that were categorized as "upright".
- Direct observation review confirmed that cycling modalities were highly variable, providing a valuable stress-test for algorithm generalizability in a sample independent from the one in which the algorithm was developed.

Key implications

This independent evaluation provides preliminary but encouraging evidence that ActiNet can accurately detect cycling duration across diverse cycling modalities, even when evaluated under stringent real-world conditions with high-quality reference measures. False-positive MVPA misclassifications highlight the need for refinement of wrist-based cycling models. Thigh-based ActiPAL performance emphasizes the importance of clear operational definitions (e.g. "cycling" vs "pedalling"), and the need for harmonized reference standards. Independent evaluation remains an essential step in wearable-algorithm development and the WAVES workplan directly addresses that gap. Future work plans include more data pooling (additional circa 250 hours identified to date) and further incorporation of more datasets in 2026 to support the development of a benchmark dataset and potential guidance on performance standards for algorithms.

Presentation 5: ActiveLIFE: Large-scale multigenerational accelerometry for physical activity and cycling surveillance

Presented by Dr Thijs Eijsvogels

Overview

This presentation introduced ActiveLIFE, a large, multigenerational population cohort in the Kingdom of the Netherlands that is integrating objective accelerometry into ongoing health surveillance. The initiative aims to link thigh and wrist accelerometer data with biological, clinical, environmental and behavioural measurements. These include electrocardiograms, blood and urine biomarkers, pulmonary function (subset), body composition, mental health, genetics (subset), CT scans (subset), risk factors, health-care expenditure, diet, environmental characteristics and health outcomes. The ActiveLIFE Human Activity Recognition sub-study will additionally support algorithm development, external validation and deeper insights into physical activity types such as cycling – a highly prevalent behaviour in the Dutch population.

Study design

- ActiveLIFE builds on the established population-based Lifelines Biobank cohort with extensive phenotyping since 2006, now expanding with accelerometry to participants aged 8–99 years.
- The ActiveLIFE cohort comprises 60 000 participants in the 4th wave of Lifelines data collection (2025–2029, n=80,000) and thus far an 87% participation rate.
- The study uses an Axivity AX3, with wear locations on right thigh (all) and non-dominant wrist (subsample of around 10%, n=6,000); sampling at 100 Hz and range: ±8 g; wear protocol: 8 days free-living.

Work underway (Human Activity Recognition sub-study)

- Currently developing a comprehensive data-processing and machine-learning framework using a development dataset of 100 participants, stratified by age (30–70 years) and cardiovascular disease (CVD) status (50% with CVD).
- These participants wear devices at six body locations and complete a free-living protocol. Activities are video annotated to produce high-quality ground truth.
- Machine-learning models are being trained and evaluated for a broad suite of behaviours (physical activity, sedentary time, steps) and specifically cycling, taking advantage of thigh-based signal quality.

Preliminary findings

- Analysis of 3555 participants with processed data showed cycling is highly prevalent with median cycling time of 58 minutes per week (inter quartile range –147 min/week).
- Plans for cycling-specific analyses include comparing wrist versus thigh outputs, different algorithmic approaches, and external versus internal validation pipelines.

Key implications

ActiveLIFE represents a new resource for objective surveillance of physical activity and cycling. Its strengths include large-scale and multigenerational sampling, harmonized wrist-thigh accelerometry, comprehensive health and environmental covariates, and purpose-built datasets for algorithm development and validation. ActiveLIFE can directly support efforts to develop robust, validated, and globally comparable algorithms – particularly for behaviours such as cycling, where wrist-only detection remains challenging. This resource will also help define reference patterns, improve existing models and inform international guidance on wearable-device based population surveillance. External validation of available algorithms is a planned study objective and can support developing a benchmarking process for algorithms.

Additional short presentation

Opportunities to advance knowledge on the specific issues of measuring cycling and more broadly on evaluation of wearable device-based measurement of physical activity was shared with an additional invited short presentation by Dr Antje Hebestreit and Dr Christoph Buck, who joined the session virtually from the University of Leibniz Institute for Prevention Research and Epidemiology (a WHO Collaborating Centre for Obesity Prevention, Nutrition and Physical Activity). They provided a brief overview of the WEALTH study[14] and shared preliminary results that indicated that wrist-worn devices were able detect cycling with 91% accuracy.

Summary of discussion

- **Classifying both activity type and activity intensity:** Classifying cycling (an activity type) is more complex than classifying intensity and to date most algorithms do not simultaneously classify both activity type and intensity; separate provision of estimates of overall MVPA and cycling time will require training datasets with both aspects and robust ground-truth labelling.
- **Detecting E-bikes and other cycling modes:** The rise of e-bikes has introduced new challenges as the movement patterns and intensity will differ from traditional cycling, and current analyses may not adequately distinguish between them. Understanding of this issue is limited by the absence of sufficient inclusion of diverse cycling modalities in training datasets and requires robust ground-truth labelling. A systematic review and the Compendium of Physical Activities for e-bikes indicate energy costs are lower when riding an e-bike compared to cycling but still meet the 3 MET threshold for moderate intensity activity. Populations interested in distinguishing manual cycling from e-bike usage may need to incorporate a self-report question on cycling modalities to supplement device data.
- **Differentiating cycling from other "similar" movement patterns:** This key issue received considerable attention and the discussion focussed on the scope and scale of potential misclassification due to activities that mimic cycling movements (e.g. pushing a stroller, driving a vehicle). It was agreed that the absence of a sufficient quantity of these activities in the training datasets is a limitation at this time; it was noted that "easy" comparisons (i.e. compared with driving a vehicle) had more data and show that current analyses can differentiate well; but it was agreed that more research was needed on the "harder" comparisons, such as differentiating cycling from riding a scooter and other activities that occur at a similar speed and have wrists placed on a handlebar.
- **How generalizable are the current (machine-learning) analytical models?** It was agreed that performance varies by dataset and context and that most validation in this field has occurred with data from high-income settings; it was therefore strongly endorsed that more evaluation studies in low- and middle-income countries are needed.

[14] See https://wealth-stamify.com/.

- **Need for reference standards for data labelling:** delegates reaffirmed the need for investment in more research to support the development of high quality more diverse training datasets to accelerate the improvement in algorithm accuracy – that is to improve detection and classification of cycling as well as MVPA .It was agreed that efforts to share and, where possible, pool datasets are needed, and harmonization of data labelling methods is a key priority. The development of common coding is needed as is establishing consensus on performance standards.
- **Leverage learning from cohort studies**: The session included presentations of secondary analyses of relevant data from large cohort studies, demonstrating the value of leveraging these large datasets for analysis to help resolve issues. The use of existing cohort datasets to inform test-specific analytic approaches was seen as potentially useful and to be pursued where relevant. Within this context it was noted that support and funding for these (expensive) longitudinal studies is hard to secure and therefore any additional dataset uses could provide a useful rationale for investment and demonstrate added value – which is vital in discussions with decision-makers and funders.
- **Is wearable device data "better" than self-report recall of cycling?** Discussion points reminded that while the focus of this session was to compare the accuracy of wrist- and thigh-worn devices, it was suggested that both wear locations are better in comparison to estimates based on self-report recall (the current method in use in most national surveillance). Counter views shared that self-report data may be better because some specific activities, such as cycling for transport to work/school, may be fairly accurately recalled and therefore provide highly valuable data. Participants agreed that this depended on survey questions. The possible use of both wearable devices *and* a questionnaire as complimentary tools was clearly stated.
- **Machine-learning methods are complex and rapidly changing:** The discussion noted that the considerable technical specificity and complexity in machine-learning methods, and that even those in the field have found it hard to keep up with large developments and new approaches over the past decade. It was also noted that this complexity adds to the challenge of communicating the evidence – within the field as well as to decision-makers. Building trust and understanding is a necessary additional agenda.

Section summary

The discussion acknowledged that accurate detection of cycling remains a critical and unresolved issue for wearable device-based physical activity surveillance, particularly in countries where cycling contributes substantially to total moderate-intensity activity. Recent advances in machine learning – especially self-supervised and GAN-based approaches – show promise for improving wrist-based cycling detection, but performance remains highly context-dependent and sensitive to training data, reference standards and cycling modality.

Delegates agreed that classifying activity type (such as cycling) requires a different methodological approach compared to estimating activity intensity, which is a physiologic

construct. Most current algorithms do not yet robustly combine both, and as a result, surveillance systems may need to report cycling time and MVPA separately. Future modelling advances explicitly trained using high-quality ground truth data with both activity type and intensity may drive advances in this area. Importantly, for surveillance, algorithm performance should be evaluated based on its impact on population-level prevalence estimates, recognizing that misclassification may influence national estimates differently than individual-level accuracy metrics suggest.

Key challenges identified included the rise of e-bikes, misclassification with cycling-like activities, and the limited generalizability of existing models, which are largely based on data from high-income settings. Participants strongly endorsed the need for more diverse training datasets and additional evaluation studies, particularly in low- and middle-income countries, and reflecting real-world cycling behaviours. A central theme of the discussion of results was the trade-off between thigh-worn devices, which logically can capture additional details about cycling such as pedalling and are more widely used at this time; and wrist-worn devices as the most scalable option for population surveillance, highlighting the importance of bridging approaches such as signal harmonization. Leveraging large cohort studies and collaborative initiatives (e.g. PRoPAss, ActiveLIFE and WAVES) for developmental research was seen as essential to accelerating progress.

The discussion reflected differing views on the relative value of wearable device data and self-report measures. While devices offer advantages for objectivity and pattern detection, self-report was considered valuable as well. For example, self-report has been shown to be accurate for measuring trips and just slightly overestimates duration. It can also provide the additional advantages of capturing other related data such as domain (e.g. exercise vs transport related cycling) and type of bicycle (e-bike, road, stationary) that might inform energy cost or other derived estimates. Finally, delegates noted that the rapid evolution and complexity of machine-learning methods pose challenges for communication and trust. Delegates suggested WHO guidance should help to translate technical findings of algorithm performance into policy-relevant terms and ensure analytic tools are transparent, accurate and reproducible. Delegates also restated the importance of independent evaluation as a prerequisite for building confidence in the use of wearable devices to measure cycling within national surveillance systems.

Section 5: Review of WHO STEPS and protocols for implementing wearable technology to measure physical activity

This section summarizes a "question and answer" session conducted with Leanne Riley, Head of the Surveillance, Monitoring and Reporting unit at WHO headquarters. Her team provides support to Member States to plan, undertake and analyse population-wide surveillance of NCD risk factors. This report provides details on the WHO STEPwise approach to surveillance (STEPS)[15] process in countries followed by a summary of the discussion on how a wearable device (wrist-worn) could be integrated into STEPS protocols effectively and efficiently. The discussion generated a list of recommended actions and support that would be required.

Introduction to STEPS and the global context

- STEPS has been implemented in approximately 120–130 countries, initially this was in low- and middle-income settings, but more recently an increasing number of high-income countries (for example in the Eastern Mediterranean Region) have also participated.
- The European Region remains less represented, as there are various pan-European surveillance systems; however uptake is starting in Central and Eastern Europe.
- STEPS surveys are recommended every 5 years but generally repeated every 7–10 years.
- Countries initiate interest and engagement through WHO regional or country offices.

Resource mobilization and WHO's role

- WHO is not a funding agency; countries are expected to mobilize their own resources to ensure sustainability.
- WHO provides technical advice, training, equipment, and support in identifying potential donor contributions where gaps exist.

[15] See https://www.who.int/teams/noncommunicable-diseases/surveillance/systems-tools/steps/instrument.

- Approximate costs of implementing one national data collection using STEPS methods and protocols range between US$ 350 000–500 000, with countries contributing staff and WHO assisting with equipment, supplies, and per diems. This excludes any use or integration of a wearable device.

Survey design and sampling

- STEPS comprises three components: (1) a questionnaire (behavioural risk factors); (2) physical measurements (height, weight, blood pressure, waist/hip circumference); and (3) biochemical assessments (blood glucose, cholesterol, urinary sodium).
- The standard sample size is around 6000 adults aged 18–69 years, stratified by sex and age; and often urban/rural at country request.
- Sampling frames are developed with relevant national authorities and based on national census data, with enumeration areas selected proportionally to urban/rural distribution.
- Only one eligible adult per household is randomly selected using electronic devices.

Field operations

- A typical field team comprises five members: driver, supervisor, two interviewers, and one interviewer/biomedical measurement technician.
- Around 15 teams operate simultaneously, each covering approx. 15–20 households per enumeration area; however these data collection protocols and capacity are determined on a country-by-country basis according to funding and local context.
- STEPS behaviour survey interviews last between 20–50 minutes; followed by physical measures undertaken in the household. Biochemical assessments are conducted at local clinics or community facilities in the same location on the following day, as they require overnight fasting and urine collection.
- STEPS usually achieves very high response rates (for example: 70–80% for Steps 1–2; 60–70% for Step 3), this is thought to be the result of strong support by government, strong community trust, and because often in low- and middle-income countries, participants have poor access to health services and participation in STEPS can provide the individual with health measures (such as blood pressure) and where needed, data collectors may also help with follow-up actions for any abnormal findings.

Training and quality assurance

- WHO conducts five days of intensive training with national data-collection teams, covering questionnaires, anthropometry, biochemical measures, and data management.
- Training includes practice sessions and pilot testing with members of the local community.

- Data are collected electronically, uploaded to secure servers, and monitored in real time for quality assurance (e.g. blood pressure intervals, household listings).
- Supervisors and WHO maintain weekly oversight calls during fieldwork; and there is strict ongoing quality assurance through data collection to detect non-compliance with data collection protocols or any other issues.

Data management and ownership

- All data are owned by the country, even when WHO provides technical or financial support.
- WHO assists with data processing, weighting, and analysis using standard packages, when requested by countries.
- Countries are encouraged to publish results promptly and anonymized datasets are deposited in the WHO Microdata Repository[16] within two years of collection.

Ethical considerations

- Informed consent to participate in STEPS is obtained at household level, with information sheets provided and revisits arranged where necessary.
- Participants may opt to complete only certain components, but this is increasingly less common and partial data are accepted.
- High participation is partly attributed to perceived health benefits and trust in government-linked survey teams.

Experience in the WHO European Region of piloting wearable devices in national surveys (Republic of Moldova and Georgia)

- Feasibility mainly depends on context and implementation. The first WHO European Region pilot (in the Republic of Moldova) demonstrated that introducing accelerometry into STEPS-style surveys is technically possible but operationally fragile when implemented under time pressure, limited training, and pandemic-related constraints.
- Training, preparation, and workflow design are critical determinants of success. In Moldova, limited training, lack of detailed guidance and positioning accelerometers as "optional" resulted in low participation (22%) and very limited usable data. In contrast, Georgia invested heavily in detailed training materials, pilot testing and redesigned workflows, leading to high wear-time compliance.
- Fieldwork conditions strongly influence data quality. COVID-19 disruptions, interviewer fatigue and competing survey demands undermined implementation in Moldova. In

[16] See https://extranet.who.int/ncdsmicrodata/index.php/home.

Georgia, motivated teams, strong coordination, and alignment with an existing salt survey enabled smoother integration of accelerometry.

- Centralized device management improves data quality. Activation of devices centrally (rather than in the field) and clearer data-handling procedures in Georgia reduced errors and improved monitoring compared to tablet-based activation used in the Republic of Moldova.
- Participant engagement strategies matter, but simple solutions may suffice. High compliance in Georgia suggests that well-designed protocols and trusted interviewers are more important than technological add-ons; SMS reminders did not significantly improve wear-time compliance.
- Iterative piloting is essential before scale-up. Across both pilots, WHO experience reinforced the need to adapt, pilot, refine and re-pilot protocols locally, ensure sufficient capacity and time, and avoid rapid deployment before workflows are robust – particularly when aiming for population-representative samples beyond capital cities.

Section summary

The discussion highlighted the operational complexity of conducting STEPS surveys, emphasizing the importance of national ownership, WHO technical support, robust training, and continuous quality assurance. Experience from WHO European Region pilots demonstrated that while integration of wearable devices into STEPS is technically feasible, success is highly dependent on preparation, workflow design and field implementation. High response rates and data quality are achievable when wearable protocols are well integrated into existing survey operations, supported by adequate training and clear procedures. Participants emphasized that integration of wearable devices requires developing realistic, feasible protocols that are piloted and adapted to country context before scale-up. The requirement for 7-day wear time was identified as one of the most critical practical challenges to implementation within STEPS protocols, and solutions are needed that do not increase cost and complexity.

Section 6: Advancing the integration of wearable devices into STEPS surveys

Given the clear outline of the STEPS protocols (see section 5) and the detailed account of the approaches taken to support STEPS implementation in Member States, this section summarizes the discussion on how the logistical challenges of practical administration could be addressed feasibly and efficiently – issues that were not addressed in detail during the 3rd meeting in Montreal *(4)*.

The discussion focused on the operational details of administering a wearable device in fieldwork; minimum data requirements (i.e., wear time, which is related to retrieval protocol); and the systems and capacities needed to support implementation the "pipeline" of wearable device use in STEPS and/or other national surveillance systems. Resolving these issues through pilot-testing and adaptation was seen as key to establish a feasible yet adaptable approach to integrating wearable accelerometer devices into the STEPS approach to national population health surveillance.

Operational considerations: a feasible mechanism for device administration

Participants were reminded that integrating accelerometers into STEPS requires alignment with existing fieldwork structures. STEPS teams typically remain in a single enumeration area for approximately 5–7 days before moving to the next location. Under current practices, a 7-day wear-time protocol introduces significant challenges, as participants often complete the required wear period after survey teams have left the area. Device retrieval was therefore identified as the primary logistical barrier and increased costs and complexity. The use of the postal service was seen to be plausible in some contexts but not suitable in many others; these concerns were further justified by experiences shared by presenting countries at this meeting. Thus, alternative methods are required and several options for device return were proposed, including:

- **Collection through local intermediaries and/or centralized community drop-off points** (such as clinics or survey sites used for blood and urine collection), supported by either a local community health worker (not a STEPS field worker) or other identified (if required, paid) local liaison person who already supports STEPS household recruitment.

- **Dedicated retrieval mechanisms**, such as specific drivers (not STEPS-related) or a local courier service, although limited availability or reliability may be a constraint.

Using a commercial courier service was considered (even though this would increase costs) and could be tested as part of future pilot studies. Furthermore, where resources allow, a courier pick up from the home or a community drop-off location could be considered, with devices delivered to a central data handling location in a major centre or capital city. The "drop-off" approach has been used elsewhere (e.g. Canada) and could be simplified by the participant returning the device to a storage box which, once all devices are registered as returned, could be transferred to a central data handling location. This approach would require enough devices to support the larger system of data collection – an issue discussed below.

Increase the number of devices through availability of affordable (low-cost) devices

A second game-changer identified to be "transformative" to adoption was access to larger numbers of more affordable accelerometers. The most immediate impact would be to reduce the urgency of speed to download data and prepare to "recycle" the devices for use by subsequent participants – often the next day and/or next community. This would ease operational pressures and allow a longer time period for device retrieval using mechanisms such as those above.

Reduce the wear-time protocol

A protocol that required fewer days of wear time might help reduce the complexity of STEPS field workers having to retrieve devices while they are still in the community. This was also regarded as potentially "transformative" for feasibility of administration in a typical STEPS field workflow. However, this raised questions about the level of evidence on how many valid days were needed to compute a population-level estimate (for example on meeting MVPA guidelines) with an acceptable level of confidence. Delegates shared knowledge on the available evidence that might support a wear-time protocol of as little as 1–3 days of wear time. The trade-offs between reducing wear time to fit the STEPS field workflow and getting sufficient quality of useable data were discussed.

One significant discussion concerned the risk of using a shorter wear time, leading to an increase in bias and uncertainty. For example, bias in the sample could be increased if the protocol only enabled participants from the first days of the STEPS fieldwork data collection cycle to be included and excluded participants interviewed on later days. These participants may have been harder to reach for the interview and have different social, economic or health factors. This possibility would increase uncertainty, potentially necessitating larger sample sizes and increased survey costs to maintain statistical precision. The discussion emphasized that methodological decisions on the number of valid wear days must be made based on population-level performance under real STEPS operational constraints.

This discussion remained inconclusive and therefore suggestions were provided on additional research and secondary data analysis that could support decision-making. These included conducting secondary analysis of existing datasets, such as UK Biobank or the ProPASS (Prospective Physical Activity, Sitting and Sleep) Consortium, other known cohort studies, and/or other national monitoring datasets held by countries represented at this meeting. Analyses should be conducted to assess the impact on population estimates of the three key metrics (i.e., MVPA, step count and sedentary behaviour) and compare and assess the following:

- 7-day protocol compared to those using shorter periods (e.g. 1, 2, or 3 days).
- Implications of not including data collected over a weekend.
- Systematic omission of data from participants who are interviewed on latter days of the data collection protocol (i.e. compare population estimates if wearable data are included only from participants interviewed on days 1–2-3 and not days 4–5-6).
- Day-of-week variability.
- Reactivity on initial wear days within studies using different wear-day protocols.

Where possible, all of these dataset issues should reflect experience in low- and middle-income settings.

While exploiting existing datasets is expedient, several issues must be considered. For example, analysis should be conducted using data collected from studies using different wear-time protocols (i.e., not only 7 days) as results from studies using a shorter than 7-day wear time might produce different results. In addition, conducting primary studies, particularly in low- and middle-income countries, may be required and provide additional knowledge generation. Moreover, it was recommended that a review of relevant published evidence was conducted to collate existing knowledge on unresolved issues.

Other technological solutions

Use of mobile phones for data transfer (e.g. Bluetooth upload) was considered but deemed impractical in STEPS and many low- and middle-income country settings because of variable access to smartphones, Internet connectivity and data plans.

Section summary

The discussion highlighted that integrating wearable devices into STEPS surveys is feasible but requires careful adaptation to existing fieldwork structures and operational realities. The primary logistical challenge identified was device retrieval under a 7-day wear-time protocol, as STEPS teams typically leave enumeration areas before participants complete wear. Alternative retrieval mechanisms – such as community drop-off points, local intermediaries, or dedicated retrieval services – were viewed as necessary in many contexts, as reliance on postal services is often impractical or unreliable.

Participants identified two potentially transformative enablers for feasibility: access to larger numbers of affordable devices, which would reduce pressure for rapid turnaround; and shorter wear-time protocols, which could better align with STEPS field workflows. However, reducing wear time raised important methodological concerns regarding bias, uncertainty, and representativeness, particularly if certain participant groups are systematically excluded. The group agreed that decisions on minimum wear-time requirements must be guided by evidence on population-level performance under real-world STEPS conditions.

To inform future guidance, participants recommended secondary analyses of existing datasets and, where necessary, new primary studies – particularly in low- and middle-income settings – to assess the impact of shorter wear protocols on key population-level metrics (i.e. MVPA, step count and sedentary behaviour). The discussion reinforced that pilot testing, staged implementation and evidence-informed adaptation are essential to establishing a scalable and context-appropriate approach to integrating wearable devices into STEPS and other national surveillance systems.

Section 7: Actions to support advancing the integration of wearable devices in WHO STEPS NCD surveillance protocols

Beyond the discussion and recommendations reported in Section 6, delegates developed a list of tasks, resources, collaborations and/or systems that might be needed by WHO or Member States to facilitate testing and adoption of wearable devices. These suggestions are summarized here.

A phased workplan of multiple pilot studies was considered essential

This would require development of guidance and potentially some toolkits on methods outlining the recommended study aims and design(s). Pilot studies should be conducted in a selection of countries with a diversity of contexts, including different country-specific survey logistics, team numbers, geography, travel schedules, weekend work patterns and cultural considerations. Forthcoming Member States planning to conduct a STEPS survey in 2026–2027 could be a starting point and they could be invited and incentivized to support and participate field pilots. The options for pilot studies include:

- test implementation of the full 7-day wear protocol as well as testing protocols using shorter wear periods (e.g. 2–3 days);
- test and compare different device retrieval models (e.g. central drop-off, community health worker collection, or driver-based retrieval).

Participants highlighted there may be some potential value in seeking any learning from analogous field operations, such as multi-day environmental sampling or data collection using multiple mobile clinics that enable one team to "leapfrog" another (that is, move one step ahead). In addition, it was strongly recommended that early pilots should include experienced accelerometry specialists to provide real-time problem solving, support field teams, and ensure fidelity to protocols.

System components required for integration into STEPS

Participants identified several essential components that WHO and country partners will need to establish before broader implementation. The list is provided in Table 2 and will be used to inform WHO work, stakeholders who collaborate with WHO, and interested Member States.

Table 2. Key components to consider when planning to use device-based measures as part of STEPS

Key requirement	Description
Devices	Enough wearable devices that meet device specifications and or standards as developed by WHO (noting that WHO currently holds approximately 1000 wearable devices for use to support advancing this work agenda, although additional procurement may also be required).
End-to-end protocol	Protocol should specify procedures for: – participant introduction to the wearable device and an explanation of device use, FAQ, participant information sheets or links to supporting materials; – device assignment, tracking and retrieval process; – management process of late returns, refusals, device loss or malfunction; – instructions, as needed, for data download, transfer, security and storage; – outline of roles and responsibilities at national, regional and WHO levels; and of any supporting collaborators.
Training materials	A dedicated module on wearables added to the STEPS fieldwork staff training package, including practical exercises, mock interviews and demonstration of device handling. Training of trainers and field staff to ensure consistency and quality.
Human resources and country capacity	A designated coordination focal point within each country STEPS team. Identification of national institutions to support technical and logistical processes. According to protocol selection, additional regional/local external consultants to deliver staff training on wearable devices and provide support through implementation (especially in early pilot work plan).
Data management and analysis pipeline	A standardized WHO data-processing "pipeline" for data handling. A methodology (ideally automated) for data analysis support to produce agreed outputs determined to be suitable for national reporting.
Quality assurance and process evaluation	Integration of accelerometer-specific quality assurance procedures into STEPS' existing quality assurance mechanisms – ideally supported by a methodology (ideally automated) for analysis of data for wear-time assessment, non-wear detection, summary metric generation (as part of an ongoing quality-control process) and immediate feedback to field teams.

Key requirement	Description
Ongoing capture of learning and dissemination	Systematic documentation of pilot experiences, including device return rates, wear compliance, logistical failures and necessary adaptations. Mechanism to share learning between pilots and future STEPS countries (website, webinars, newsletter, community of practice).
Advocacy and communication	Develop communication tools to provide clear messaging and justification for ministries of health, education and/or sport or other entity including potential financial supporters, highlighting the added value of device-based measurement and its policy relevance. Share examples from pilot studies and other countries to demonstrate the value of wearable device data and its use both with and instead of self-reported data on physical activity.

Section summary

The discussion provided delegates with detailed insights into STEPS protocols and examples of how these have been implemented and adapted to country contexts. Participants identified a set of priority actions to develop guidance and protocols on wearable devices to better align with existing STEPS field workflow. A phased workplan for pilot-testing was outlined and strongly recommended.

Many participants expressed a strong willingness to support WHO work, including leveraging opportunities to conduct pilot studies and contributing to the development of training materials, tools and other resources required for accelerometer-based measurement in STEPS. A list of the resources that were considered essential or desirable was generated to guide WHO planning pilot and demonstration studies. There was broad agreement that further pilot testing is necessary before wearable devices can be more widely recommended for routine implementation at scale within STEPS or other national health surveillance systems.

Section 8: Global research: current and future opportunities

The meeting discussed sharing of information on current collaborations and opportunities to collectively advance the evidence, technical and operational implementation, and policy related to the use of wearable devices to measure physical activity and related behaviours in national population surveillance systems. The session was supported by four presentations followed by plenary discussion.

Presentation 1: WAVES research initiative (Wrist Algorithms: Verification, Evaluation and Surveillance)

Presented by Dr Daniel Fuller on behalf of the international research group

The WAVES initiative and workplan responds to increasing recognition that while many algorithms exist to classify core behaviours such as steps, sedentary time and MVPA, they have not yet been systematically compared or independently evaluated under real-world conditions for global surveillance purposes. WAVES research is supporting progress towards harmonized evaluation of algorithms. This requires more independent testing and coordinated technical work to address long-standing uncertainties in order to inform global, wearable-based surveillance. The WAVES initiative was developed as a direct response to the discussion and identified limitations outlined in the 3rd expert meeting in Montreal.

Purpose and scope

WAVES aims to evaluate and compare available wrist-worn accelerometer algorithms using rigorous "Phase III" performance evaluation criteria *(32)*. This includes testing algorithms in real-world settings, using established reference measures ("ground truth"), diverse participant samples, and appropriate statistical methods. A core objective is to identify a set of valid, reliable, transparent and globally applicable algorithms to compute population estimates of step counts, sedentary time and (critically) MVPA to support the national surveillance work of countries and WHO.

Study 1: Algorithm evaluation using video-labelled data

The first component focuses on identifying existing datasets that include video-labelled free-living behaviour alongside wrist-worn accelerometer data. Several suitable datasets have been identified across different populations. A total of six step algorithms, eight

MVPA algorithms, and six sedentary behaviour algorithms have been shortlisted for head-to-head comparison. Current work includes harmonizing datasets, creating a shared development environment, and establishing open-source code repositories.

Study 2: Field-based evaluation

The second component will conduct a field-based assessment using existing datasets that contain wrist accelerometry paired with a secondary reference device that pragmatically can be worn by many participants, such as a thigh-worn accelerometers for sedentary time *(14)*. This study aims to complement the video-based work and inform the protocol and algorithm selection for a future multi-country validation study, with a focus on low- and middle-income settings.

Next steps

Future plans are to develop robust benchmark datasets, as well as transparent and shared methods for others to use to ensure repeatability and scalability. WAVES research work commenced in 2025 and is ongoing and aims to develop guidance on secondary reference measures, and to recruit additional datasets and collaborators. It was noted that whilst significant technical expertise exists globally, financial and institutional support for coordination and funding activities will be essential to maintain and accelerate progress. In due course, WAVES will invite contributions of additional benchmark datasets and participation from the broader research community.

Presentation 2: Embedding wearables into existing low- and middle-income cohorts

Presented by Dr Ben Maylor on behalf of the University of Oxford Wearables Group, and collaborators

The presentation outlined plans to embed wearable wrist-worn accelerometry into three large, ongoing cohort studies in low- and middle-income countries. These activities will be implemented over during 2026 and are designed to assess feasibility, strengthen methodological capacity, and generate behavioural data from rural populations that are typically under-represented in global datasets.

Study 1. Agincourt Health and Demographic Surveillance System (South Africa)

The Agincourt cohort follows approximately 115 000 individuals across 31 rural villages, with long-standing surveillance of demographic and chronic disease outcomes *(33)*. Ethics approval has been secured to incorporate wearable wrist-worn devices for the first time. A feasibility study will soon start, with device provision and technical oversight led by the research team. This work will support the development of protocols suitable for rural African contexts and inform future scale-up.

Study 2. India Study of Healthy Ageing (ISHA)

ISHA is a large national cohort that recruits a representative subsample of approximately 10 000 adults every 5–10 years from an underlying population of about 220 000 *(34)*. A dual-energy x-ray absorptiometry (DEXA) scan substudy (around 8000 participants) is currently under way. Ethical approval has been granted to embed wrist-worn wearables into the ongoing DEXA data collection, with device deployment and field-team training scheduled for early 2025. The study operates in remote rural areas, with participants transported by bus to central facilities, offering an opportunity to test wearable deployment under challenging logistical conditions.

Study 3. The Malaysia Cohort (TMC)

The Malaysia Cohort includes 106 000 adults from 95 rural and 56 urban locations. A new accelerometer sub-study (n ≈5000) has recently received ethical approval. The Oxford Wearables Group and the ProPASS consortium will support capacity building for the local data-collection team.

Planned device validation studies

For both the Agincourt and ISHA cohorts, the accelerometer sub-studies will include an additional validation component involving a smaller sample (around 50 people per site). As per WAVES guidance, these validation sub-studies will include multiple simultaneous reference measures (e.g. wearable camera, thigh-worn devices for sedentary behaviour, ankle devices for step-count validation, and heart-rate monitoring for MVPA). These validation datasets will be critical for improving the generalizability of behaviour-classification models, particularly for low- and middle-income populations whose daily routines differ substantially from those represented in existing validation datasets from the United Kingdom of Great Britain and Northern Ireland.

Linkages and governance

In response to a query during the session, the presenter clarified that these wearable and validation sub-studies are embedded within existing research cohorts, rather than national surveillance systems, and are not fully representative of national populations. For example, the Malaysian Cohort benefits from central government funding, although engagement varies across ministries and is driven primarily by scientific rather than public-health mandates.

Presentation 3: Opportunities for WHO method testing in low- and middle-income countries through the ProPASS Consortium

Presenter Dr Manos Stamatakis on behalf of ProPASS consortium

The presentation introduced opportunities to advance testing of WHO-recommended methods and to support development of models in low- and middle-income countries

through the ProPASS Consortium. ProPASS is an international network comprising more than 45 studies using wrist- or thigh-worn accelerometry. Unlike traditional consortia, ProPASS not only aggregates data but also actively supports new studies – particularly in under-represented regions – to embed accelerometry in ongoing cohort work. This positions the consortium as a valuable platform for WHO's aims to generate diverse, high-quality data for algorithm training and field-testing.

Several near-term opportunities were highlighted. ProPASS has recently completed collaborative agreements with a Healthy Workers Cohort in Mexico, which plans to begin wrist-worn accelerometry with approximately 2000 participants in early 2026. Although administrative processes were lengthy, collaboration is now well-established and could be mobilized to test WHO protocols. A second, highly promising opportunity lies in the Malaysia Cohort, already partnering with both ProPASS and the Oxford team, with strong local enthusiasm to incorporate accelerometry and good potential for both field-testing and algorithm development. Additional possibilities were mentioned in Thailand, where prior testing has been undertaken with thigh-worn sensors, offering a potential entry point for WHO-aligned methodological work.

The importance of incorporating validation subsamples within these low- and middle-income cohorts – ideally including multiple simultaneous reference measures (e.g. thigh-worn devices for sedentary behaviour, ankle devices for step-count validation, and heart-rate monitoring for MVPA) – was emphasized. Such datasets would support the development of more generalizable behaviour-classification models, complementing work within WAVES. Finally, it was emphasized to frame these opportunities and engagements as mutually beneficial academic collaborations. This should include incentives around, for example, co-authorship in joint publications and shared outputs to encourage participation and sustain long-term partnerships in low- and middle-income research settings.

Presentation 4: Wearable landscape analysis

Presenter Dr Marco Giurgiu

The presentation described ongoing work to support WHO's development of technical guidance on the use of wearable devices in national population surveillance of physical activity. As part of this process, WHO has commissioned a comprehensive *landscape analysis* to identify which currently available wearable devices meet the emerging device specifications. This work aims to provide WHO and Member States with a clear, evidence-based overview of device options suitable for use in large-scale surveillance.

The landscape analysis uses a systematic review and AI-assisted Internet search strategy to identify devices and has a structured protocol (building on a prior systematic review that covered the period 1970–2022) and identified 423 devices. Of these, only 48 remain commercially available. Additional searches identified a further 278 devices, resulting in a consolidated catalogue of 326 wearable devices across 101 companies. For each device, an extensive set of characteristics is being extracted from publicly available materials – this

covers details such as sensor types, wear locations, data export capabilities, battery life, and feedback functions – using a structured data-collection framework.

Outputs of this landscape analysis include: a complete device database; detailed analyses for selected devices; an evaluation report; and an online, searchable dashboard that will be publicly accessible and updated annually. The dashboard will enable users to search for devices, view their specifications, and access links to validation studies.

The database can be filtered according to the device specifications emerging from recommendations from WHO expert meetings. For example, of the 326 wearable devices identified, 220 provide behavioural feedback, which is unsuitable for population surveillance, thus narrowing the pool to 106 devices. Applying further technical filters – such as presence of a triaxial accelerometers and raw data export – reduced this set to approximately 51 potentially relevant devices. The team expects that further refinement will produce a shortlist of 20–30 devices that will then be subject to deeper analysis. This "living review" will serve both WHO and the wider research community as a resource for identifying devices that align with forthcoming WHO needs and recommendations. More information on this work as well as other resources and access to an on-line data base developed by the Wearable Landscape Consortium is available at https://wearable-landscape.info/network.

Summary of discussion

1. Expanding participation and regional engagement

- Participants highlighted the need for broader geographical engagement in this field, particularly in regions with high interest but limited representation at the meeting (e.g. Western Pacific, PAHO/Latin America, parts of Eastern Europe).
- WHO regional offices have begun identifying some candidate countries for future involvement – including El Salvador, Zambia, and others – that may be willing to pilot-test wearable-based measurement.
- WHO suggested convening additional regional-focused webinars and informational meetings as a strategy for building interest, consolidating partnerships and identifying new pilot countries.

2. Leveraging knowledge and capacity from existing networks and facilitating global collaboration

- The value of leveraging existing research networks (e.g. WAVES, ProPASS and other collaborations not represented at the meeting) was emphasised as a mechanism for connecting motivated countries and building technical expertise
- Countries expressing interest in beginning accelerometry work should consider linking directly with researchers already active in their region (e.g. Mexico with ProPASS collaborators; South Africa with Oxford Wearables Group).

- There is strong goodwill and capacity within these networks to develop early and mid-stage research professionals and support early-stage pilots, including by training and technical troubleshooting.

3. Deliverables and expected outcomes from the WAVES research initiative

- WAVES expects to provide methods for and results from independently tested algorithms, thereby providing WHO with options to include in the WHO guidance document on how to calculate step count, sedentary behaviour and MVPA measurements from device data.
- Participants noted that methods and algorithms for step count and sedentary classification are nearly ready for use, though MVPA requires further work and additional training data, especially from older adults and low- and middle-income countries, and for some behaviours of interest (e.g. cycling).
- A primary output from WAVES will be the creation of independent "benchmark" datasets that allow transparent, reproducible evaluation of any algorithm against standardized criteria.

4. Need for standards and accuracy thresholds

- Discussion highlighted the absence of agreed benchmarks for what constitutes "acceptable" algorithm performance and establishing such standards would guide both algorithm evaluation and country-level adoption.
- Participants proposed developing expert-agreed minimum standards (e.g. accuracy, sensitivity, specificity thresholds) informed by practice in other fields.

5. Combining pilot "demonstration" studies with evaluation sub-studies

- Several participants emphasized the opportunity and value of pilot (or "demonstration") studies including a small validation sub-study, where this sub-sample includes a ground-truth reference measure (e.g. camera-assisted annotation, thigh devices for sedentary behaviour, ankle devices for step count, heart rate for MVPA). This is an efficient way to advance building diversity into the training datasets (see point 3).
- Collecting such data in low- and middle-income countries, or rural settings, is feasible and would substantially improve generalizability of models, especially for under-represented movement behaviours (such as cycling, and other country and context specific occupational activities).

6. Practical and operational learning from existing multi-country cohort studies

- It was recommended that the experience from epidemiological studies underway in low- and middle-income countries, including in rural areas (e.g. in Malaysia, Mexico, and South Africa), would help identify common operational issues, such as downloading data with limited connectivity, device troubleshooting and participant adherence.

- Participants noted that WHO field staff during STEPS may find it difficult to manage complex data handling of wearable datasets, meaning third-party technical support structures will be required to assist ministries during early adoption.

7. Implications for future WHO support

- WHO's role is expected to centre on facilitating networks, providing methodological options, and coordinating third-party technical assistance for countries.
- The development of a "pipeline" of support to countries to address data handling and processing was identified as essential for future integration of wearables into STEPS at scale.

8. Wearable device landscape, technical product profile and repository sustainability

- WHO highlighted the commencement of a market landscape analysis that draws on substantial prior investment by collaborating research groups. This repository of knowledge will directly inform development of WHO guidance by identifying which, if any, commercially available devices meet the emerging design specifications.
- Questions were raised about the long-term sustainability of the repository, including the resources required for maintenance, annual updates, and data verification.
- The research lead confirmed plans for the repository and open access data platform to evolve into a collaborative, user-contributable resource, contingent on future funding, with annual updates envisaged as part of a "living review" model.

Section 9: Data sharing: an example of a national platform

This session focused on national data infrastructure as a critical enabler of wearable-based measurement and evidence-informed policy-making in sport and physical activity. As countries seek to integrate wearable data with other health and population surveillance data and policy systems, practical models for secure, governed and interoperable data sharing will be increasingly important. The Dutch Sports Data Valley initiative was presented as a real-world example of how such an ecosystem can operate at national scale.

Sports Data Valley, Netherlands (Kingdom of the)

Presented by Dr Remco Hoekman and Dennis van Kooij

This presentation introduced Sports Data Valley (SDV) as the Dutch national data infrastructure designed to support research, policy, education and practice in sport, physical activity and health.

Established and led by the Mulier Institute, SDV was developed in response to the growing need for a coherent national data infrastructure for sport and physical activity. The platform seeks to address current fragmentation in data collection, storage and access, and strengthen evidence-based decision-making through secure, privacy-compliant and interoperable data sharing.

SDV is an integrated platform and brings together: questionnaire, registry and survey data; wearable and sensor data, including from major commercial wearables with established data connections; tools for analysis, reporting and secure data storage within the Kingdom of the Netherlands. It is an open, publicly funded infrastructure accessible to citizens, researchers, trainers, and policy-makers. The platform shows how data from wearables, questionnaires and facilitates can be stored and provide secure sharing and re-use of both individual and population-level data.

All data are hosted within the country to ensure compliance with privacy and ethical standards. Users retain ownership and full control through transparent access logs, informed consent, and individualized sharing rules, allowing participants and research teams to determine how and with whom data are shared. The platform supports a wide range of applications including the following:

- **Research**: enables longitudinal monitoring, integration of multiple data types and real-time access for approved researchers, including studies of athlete training load, multimodal monitoring and clinical applications using consumer devices with electrocardiography functionality.
- **Education**: allows students to gain practical skills in data collection, management and analysis.
- **Clinical monitoring** of patients such as those with cardiomyopathy using daily wearable data.
- **Elite athlete monitoring:** supporting coaches and athletes monitor their well-being and training load, with the possibility of combining wearable and questionnaire data.
- **Policy**: offers interactive dashboards summarizing key indicators at national, regional and local levels to inform planning, evaluation and monitoring.

SDV represents a national collaboration across research, policy and practice, bringing disparate data sources together under a trusted governance model to strengthen multisectoral collaboration, improve the quality and availability of evidence, and support more coherent, informed and transparent policy-making in sport, physical activity and health.

SDV has encountered challenges in integrating wearable data, including variability in outputs across device models, limited access to raw sensor streams, and differences in update frequency driven by device-specific application programming interfaces.

Important considerations for countries exploring wearable-based measurement were identified, including establishing trusted data governance arrangements and ensuring technical flexibility to accommodate heterogeneous data types and evolving device ecosystems.

Section 10: Recommendations: priorities and suggested next steps

The meeting reaffirmed the importance and timeliness of WHO's convening and normative role in strengthening the scientific and technical methods, and associated guidance, for the use of wearable devices to measure physical activity and sedentary behaviour within national surveillance systems. It reconfirmed strong interest and support from academic partners and national governments to collaborate, including with WHO, to advance this agenda at a critical moment, as countries seek more timely, policy-relevant and actionable data to guide prevention strategies.

While wearable devices were recognized as offering important advantages over self-report questionnaire instruments, several technical, methodological and operational issues must still be resolved before broader national implementation is feasible. Participants emphasized the need for a phased approach to adoption, through, for example, use of the "trial/test and adapt" approach. The meeting also highlighted the importance of clear, robust and policy-relevant communication resources to support engagement with decision-makers and to ensure confidence in surveillance data during periods of methodological transition.

The priority actions identified as next steps to advance this agenda are summarized below across five broad areas.

1. Completion of WHO guidance on the use of wearable devices in national population health monitoring systems

Completion of WHO guidance on the use of wearable devices in national population health monitoring systems is a core normative deliverable that underpins all other recommended actions. Finalizing this guidance will require additional analytical work and targeted consultation in several priority areas, including on additional issues identified through this meeting. This priority work will include:

a. Minimum and optimum device requirements and protocols

Undertake additional data analysis and consultation to complete drafting of minimum and optimum requirements for wearable devices in adults (and in adolescents), including secondary data analyses to inform decisions on key parameters such as wear-time criteria, data handling, and data formats for storage and sharing, as appropriate.

b. Training datasets and ground-truth standards

Develop larger and more diverse training datasets, employing common or harmonized labelling systems and recommended ground-truth reference instruments, as required. This work should seek to leverage and pool existing datasets and build on activities already underway through the WAVES research consortium.

c. Benchmarking and algorithm performance assessment

Establish a standard benchmarking dataset and a set of performance standards to support independent assessment of open-source algorithms. Develop an accessible and sustainable mechanism to support this global work on an ongoing basis, such as through a global hub or coordinated platform.

d. Data governance, ethics and regulatory alignment

Consult and develop guidance on data governance for a wearable-devices dataset on physical activity and related behaviours including ethical and regulatory considerations, consent models, cross-border data-sharing constraints (particularly where international support for data processing is required), and safeguards for the responsible reuse of raw data.

e. Integration with self-report measures

Consult and develop practical examples of different approaches and multiple use-cases for wearable device-based measures of physical activity and sedentary behaviour, including their use as a replacement for, or in combination with, questionnaire-based self-report data within national population surveillance systems, with a focus on informing national policy and practice.

The overall aim of the guidance should be to strengthen methodological approaches for countries considering testing and adopting wearable devices and it should align with the best available scientific evidence. It should support national and global comparability of device-based data while allowing flexibility and adaptation of protocols – particularly for device administration and retrieval – to accommodate diverse country contexts. The document should provide both practical and technical direction to support countries adopting or modifying protocols for wearable device use in national surveillance systems.

2. Research and development

While substantial progress has been made over the past decade, important knowledge gaps remain, particularly in relation to research on the use of wearable devices within population health surveillance systems and for tracking population-level trends. Targeted research is therefore needed to strengthen the evidence base for device-based measurement in these contexts. Efforts should seek resources to support and expand international collaborations to evaluate wearable device performance across diverse populations, especially in low- and

middle-income countries, improve algorithms (particularly for measuring MVPA), and strengthen evidence on generalizability.

Opportunities and partnerships should be pursued and supported to build capacity across academic and research institutions, as well as within government and national statistical agencies, ensuring that methodological advances translate into usable evidence for policy and planning. Scientific and technical collaborations between academic partners, governments and other stakeholders can support innovation while accelerating uptake in surveillance practice. A key equity-related evidence gap is the limited data on device performance among people with atypical or impaired movement, including those living with disabilities or altered gait, and in older adults, which must be addressed to improve algorithm accuracy and avoid systematic exclusion.

Plan and conduct experimental studies, potentially within countries participating in or preparing to conduct WHO STEPS, to test strategies to improve response rates and the representativeness of samples in device-based measurement of physical activity and sedentary behaviour. This should include evaluation of recruitment strategies for under-represented groups; different incentive and feedback approaches; and operational models for device distribution, use and return – particularly in low- and middle-income settings where postal systems are limited, including community-based and hybrid delivery models.

3. Conduct demonstration studies

National adoption should be envisaged as a phased approach, where Phase 1 involves a number of (e.g. five to seven) demonstration studies as a deliberate sequencing step between methodological development and broader national uptake. These should be planned and implemented, leveraging where possible opportunities within countries planning to conduct WHO STEPS or other national health surveillance systems, ideally during 2026–2028. Learning from these frontrunner countries representing diverse contexts should be used to refine and update WHO guidance, protocols and training tools, as required. Participating countries should be convened to share experiences, lessons learned and practical solutions, strengthening collective knowledge and readiness for subsequent phases of implementation

A mechanism should be established to provide technical support in demonstration sites and respond to country requests for assistance, including delivery of training on wearable device deployment, data management and data processing. Opportunities should be sought to deliver this support through collaboration with existing institutional structures, WHO networks and relevant research consortia. Such a mechanism will be particularly important for low- and middle-income countries, and for rural and remote contexts, where current expertise and access to technologies may be more limited.

Opportunities, partnerships and funding mechanisms should also be sought to develop and maintain a shared stock of wearable devices and supporting components (e.g. batteries, repair, maintenance and inter-country transport) for use in demonstration studies. The cost

of devices and the need for sufficient numbers of them remain key constraints that must be addressed to enable equitable adoption of technology. In parallel, national capacity and collaboration among authorities, technical experts, research groups, statistical offices and research institutes should be strengthened to build sustainable systems and infrastructure for the use of wearable technology, including training in data collection and management, processing and analysis, and use of open-source analytical methods.

4. Communications, advocacy and alignment

Communication, engagement and advocacy materials should be developed and disseminated to support informed dialogue with governments and policy-makers on the purpose, value and implications of testing, adopting and integrating wearable devices into national health surveillance systems. Policy-facing products – including concise rationales, policy briefs, technical notes and country examples – can support decision-makers in understanding the added value of device-based measurement alongside existing approaches, without undermining confidence in established surveillance systems. Guidance should also support countries in communicating and interpreting results, including explaining uncertainty and discrepancies between device-based and questionnaire-derived estimates.

In addition, actions and communications should ensure alignment between guidance on the use of wearable devices in surveillance systems and the evolving epidemiological evidence on the health benefits of physical activity and the risks associated with sedentary behaviour. The planned update of the WHO Global guidelines on physical activity and sedentary behaviour in 2030 presents a strategic opportunity to ensure that advances in device-based measurement are appropriately reflected in future guidance. This alignment will help future-proof WHO guidance on wearables and support surveillance systems to remain adaptable as evidence evolves.

5. Resource mobilization and investment

Delivering this programme of work will require sustained and coordinated investment from governments, research funders and other stakeholders with a shared interest in strengthening population health intelligence. As countries increasingly rely on data to guide policy decisions, allocate resources and assess impact, investment in robust, comparable and policy-relevant measurement systems represents essential infrastructure for effective and equitable action. Without such investment, there is a risk that advances in measurement technology will outpace the ability of health systems to use data consistently and responsibly.

Targeted support is needed to underpin the analytical work, consultations, demonstration/phase 1 studies and capacity-building activities outlined above, including investment in shared data resources, technical platforms and training. These investments will enable countries to move beyond exploratory use of wearable devices towards scalable and

sustainable integration within national surveillance systems, ensuring that improved measurement leads to better evidence, better policy decisions and greater health impact.

Opportunities should be sought to align this agenda with existing national, regional and global research and development priorities, including digital health, noncommunicable disease prevention, health system strengthening and equity-focused innovation. Strategic co-investment across governments, research agencies, philanthropic organizations and other partners can accelerate progress, reduce duplication and ensure that advances in device-based measurement benefit countries at all income levels.

References[17]

1. Global action plan on physical activity 2018-2030: More active people for a healthier world. Geneva: World Health Organization; 2018 (https://apps.who.int/iris/handle/10665/272722) Licence: CC BY-NC-SA 3.0 IGO.
2. Physical activity measurement and surveillance in adults: Report of a scoping and planning meeting, 27-28 november 2023. Geneva: World Health Organization; 2024 (https://iris.who.int/handle/10665/378157) Licence: CC BY-NC-SA 3.0 IGO.
3. Physical activity measurement and surveillance in youth: Report of a scoping and planning meeting, 8-10 May 2024. Geneva: World Health Organization; 2024 (https://iris.who.int/handle/10665/379652) Licence: CC BY-NC-SA 3.0 IGO.
4. Measuring physical activity in adults using wearable technologies: Report of a scientific meeting, 10-12 June 2025, Montreal, Canada. Geneva: World Health Organization; 2025 (https://iris.who.int/handle/10665/383363) Licence: CC BY-NC-SA 3.0 IGO.
5. Global strategy on digital health 2020-2025. Geneva: World Health Organization ; 2021 (https://iris.who.int/handle/10665/344249) Licence: CC BY-NC-SA 3.0 IGO.
6. Rowlands AV, Mirkes EM, Yates T, Clemes S, Davies M, Khunti K et al. Accelerometer-assessed physical activity in epidemiology: are monitors equivalent? Med Sci Sports Exerc. 2018;50(2):257–65 (https://www.ncbi.nlm.nih.gov/pubmed/28976493).
7. Migueles JH, Molina-Garcia P, Torres-Lopez LV, Cadenas-Sanchez C, Rowlands AV, Ebner-Priemer UW et al. Equivalency of four research-grade movement sensors to assess movement behaviors and its implications for population surveillance. Sci Rep. 2022;12(1):5525 (https://www.ncbi.nlm.nih.gov/pubmed/35365696).
8. Brocklebank L, Steptoe A, Zhao Y, Bloomberg M, Acquah A, Yuan H et al. Comparing two wrist-worn accelerometers (Axivity AX3 and Matrix 003) for measuring movement behaviours in British and Chinese older adults. J Meas Phys Behav. 2025;8(1) (https://doi.org/10.1123/jmpb.2025-0011).
9. Crowley P, Skotte J, Stamatakis E, Hamer M, Aadahl M, Stevens ML et al. Comparison of physical behavior estimates from three different thigh-worn accelerometers brands: a proof-of-concept for the prospective physical activity, sitting, and sleep consortium (PROPASS). Int J Behav Nutr Phys Act. 2019;16(1):65 (https://www.ncbi.nlm.nih.gov/pubmed/31419998).
10. Buchan DS. Equivalence of activity outcomes derived from three research grade accelerometers worn simultaneously on each wrist. J Sports Sci. 2022;40(7):797–807 (https://www.ncbi.nlm.nih.gov/pubmed/34962185).
11. Liu F, Wanigatunga AA, Schrack JA. Assessment of physical activity in adults using wrist accelerometers. Epidemiol Rev. 2022;43(1):65–93 (https://www.ncbi.nlm.nih.gov/pubmed/34215874).
12. Troiano RP, McClain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity research. Br J Sports Med. 2014;48(13):1019–23 (https://bjsm.bmj.com/content/48/13/1019).

[17] All links accessed 9 February 2026 unless otherwise indicated.

13. Giurgiu M, von Haaren-Mack B, Fiedler J, Woll S, Burchartz A, Kolb S et. al. The wearable landscape: issues pertaining to the validation of the measurement of 24-hour physical activity, sedentary, and sleep behavior assessment. J Sport Health Sci. 2024 (https://pubmed.ncbi.nlm.nih.gov/39491744/).

14. Matthews CE, Saint-Maurice P, Freeman JR, Hayes HA, Shreves AH, Doherty A et al. Performance evaluation of algorithms to estimate daily sedentary time using wrist-worn sensors in free-living adults. J Meas Phys Behav. 2025;8(1) (https://doi.org/10.1123/jmpb.2024-0051).

15. Suorsa K, Pulakka A, Leskinen T, Pentti J, Holtermann A, Heinonen OJ et al. Comparison of sedentary time between thigh-worn and wrist-worn accelerometers. J Meas Phys Behav. 2020;3(3):234–43 (https://doi.org/10.1123/jmpb.2019-0052).

16. Marcotte RT, Petrucci GJ, Jr., Cox MF, Freedson PS, Staudenmayer JW, Sirard JR. Estimating sedentary time from a hip- and wrist-worn accelerometer. Med Sci Sports Exerc. 2020;52(1):225–32 (https://pmc.ncbi.nlm.nih.gov/articles/PMC9624389/).

17. Montoye AHK, Moore RW, Bowles HR, Korycinski R, Pfeiffer KA. Reporting accelerometer methods in physical activity intervention studies: A systematic review and recommendations for authors. Br J Sports Med. 2018;52(23):1507–16 (https://doi.org/10.1136/bjsports-2015-095947).

18. Kamada M, Shiroma EJ, Harris TB, Lee IM. Comparison of physical activity assessed using hip- and wrist-worn accelerometers. Gait Posture. 2016;44:23–8 (https://www.sciencedirect.com/science/article/pii/S0966636215009418).

19. Gall N, Sun R, Smuck M. A comparison of wrist- versus hip-worn actigraph sensors for assessing physical activity in adults: A systematic review. J Meas Phys Behav. 2022;5(4):252–62 (https://doi.org/10.1123/jmpb.2021-0045).

20. Tudor-Locke C, Barreira TV, Schuna JM, Jr. Comparison of step outputs for waist and wrist accelerometer attachment sites. Med Sci Sports Exerc. 2015;47(4):839–42 (https://doi.org/10.1249/mss.0000000000000476).

21. Maylor BD, Edwardson CL, Clarke-Cornwell AM, Davies MJ, Dawkins NP, Dunstan DW et al. Physical activity assessed by wrist and thigh worn accelerometry and associations with cardiometabolic health. Sensors (Basel). 2023;23(17) (https://www.ncbi.nlm.nih.gov/pubmed/37687813).

22. Ahmadi MN, Koemel N, Biswas R, Holtermann A, Koster A, Atkin AJ, et al. Development and evaluation of wrist- and thigh-worn accelerometer algorithms using self-training machine learning models for classification of activity type and posture: Towards device placement-agnostic methods in the PROPASS consortium. preprint. 2025.

23. Koffman L, Muschelli J. Evaluating step counting algorithms on subsecond wrist-worn accelerometry: A comparison using publicly available data sets. J Meas Phys Behav. 2024;7(1) (https://doi.org/10.1123/jmpb.2024-0008).

24. Walmsley R, Chan S, Smith-Byrne K, Ramakrishnan R, Woodward M, Rahimi K et al. Reallocation of time between device-measured movement behaviours and risk of incident cardiovascular disease. Br J Sports Med. 2021;56(18):1008–17 (https://bjsm.bmj.com/content/bjsports/56/18/1008.full.pdf).

25. Ramakrishnan R, Doherty A, Smith-Byrne K, Rahimi K, Bennett D, Woodward M et al. Accelerometer measured physical activity and the incidence of cardiovascular disease: Evidence from the UK biobank cohort study. PLoS Med. 2021;18(1):e1003487 (https://pmc.ncbi.nlm.nih.gov/articles/PMC7802951/).

26. Chen Y, Chan S, Bennett D, Chen X, Wu XY, Ke Y et al. Device-measured movement behaviours in over 20,000 China Kadoorie Biobank participants. Int J Behav Nutr Phys Act. 2023;20(1):138 (https://doi.org/10.1186/s12966-023-01537-8).

27. Wolff-Hughes DL, McClain JJ, Dodd KW, Berrigan D, Troiano RP. Number of accelerometer monitoring days needed for stable group-level estimates of activity. Physiol Meas. 2016;37(9):1447–55 (https://pubmed.ncbi.nlm.nih.gov/27510765/).

28. Doherty A, Jackson D, Hammerla N, Plötz T, Olivier P, Granat MH et al. Large scale population assessment of physical activity using wrist worn accelerometers: The UK Biobank Study. PLoS One. 2017;12(2):e0169649 (https://doi.org/10.1371/journal.pone.0169649).

29. Cycling facts 2023. The Hague: Netherlands Institute for Transport Policy Analysis (KiM); 2023 (https://english.kimnet.nl/documents/2024/01/10/cycling-facts-2023).

30. Buehler R, Goel R. Chapter seven – a global overview of cycling trends. In: Heinen E, Götschi T, editors. Advances in transport policy and planning. 10: Academic Press; 2022. p. 137–58 (https://doi.org/10.1016/bs.atpp.2022.04.007).

31. Exercise guidelines [website]. Bilthoven: National Institute for Public Health and the Environment; 2024 (https://www.Sportenbewegenincijfers.Nl/kernindicatoren/beweegrichtlijnen).

32. Keadle SK, Lyden KA, Strath SJ, Staudenmayer JW, Freedson PS. A framework to evaluate devices that assess physical behavior. Exerc Sport Sci Rev. 2019;47(4):206–14 (https://www.ncbi.nlm.nih.gov/pubmed/31524786).

33. Kahn K, Collinson MA, Gómez-Olivé FX, Mokoena O, Twine R, Mee P et al. Profile: Agincourt health and socio-demographic surveillance system. Int J Epidemiol. 2012;41(4):988–1001 (https://doi.org/10.1093/ije/dys115).

34. Mhatre SS, Bragg F, Panse N, Judge PK, Manjrekar A, Burrett JA et al. Cohort profile: Indian study of healthy ageing (isha-barshi). Int J Epidemiol. 2024;53(4):dyae079 (https://doi.org/10.1093/ije/dyae079).

Annex 1: Programme

Day 1: Tuesday 25 November 2025

Time	Item	Session moderator	Location: Figi – conference room
12:00–13:00	Lunch		In-person arrival and registration at Hotel Figi Lunch and networking
13:00–13:45 20:00 Japan 12:00 UK 5.00 Arizona 4.00 California	**SESSION 1** **Welcome, introductions and aims**	WHO HQ RIVM	**Welcome** by the co-hosts of this meeting: WHO and RIVM **Set the current context and agenda at WHO** Outline background context, expectations of meeting and intended outputs **Set the current context and agenda at RIVM / NL** Short introduction to (1) physical activity in health policy and population monitoring in the Kingdom of the Netherlands and (2) what RIVM would like out of this meeting to help progress use of wearable devices for population monitoring in the Kingdom of the Netherlands. **Introductions –** allow 15–20 mins Round table of participants
13:45–14:30 21:15 Japan 12:45 UK 5.45 Arizona 4.45 California	**SESSION 2** **Overview of WHO and work on wearables at WHO HQ and European Region**	WHO HQ WHO EURO	**Presentation** 1–15 mins + Q+A WHO HQ overview of: (1) workplan to develop global guidance on the integration of wearable devices into population surveillance systems – includes summary of prior meetings, progress made, identified challenges and research needs; (2) outline planned work and contribution of this meeting; and (3) briefly introduce country completed summary and key discussion areas where input it needed most. **Presentation 2** – 15 Mins EURO experiences with including wearable devices in population health monitoring. Q&A
14.30–14.45	BREAK		

Time	Item	Session moderator	Location: Figi – conference room
14:45–16:30 22:15 Japan 13:45 UK 6.45 Arizona 5.45 California	**SESSION 3** **Learning from country experiences**	WHO HQ Speakers from countries	**Introduction to session, purpose and format** **Presentation to provide overview of country approaches (using completed WHO template)** WHO HQ – 15 minutes **Country presentations:** **1 Canada – Canadian Health Measures Survey (CHMS)** Presenters: Rachel Colley & Stephanie PrinceWare 15 minutes **2 Finland – Healthy Finland 2022–23** Presenter: Jouni Lahti 15 minutes **3 Denmark – Danish National Health Survey (NHNS) 2023 and Moving Denmark 2020** Presenter: Mette Toftager 15 minutes **4 Japan – National Health & Nutrition Survey – Japan (NHNS-J)** Presenter: Shigeru Inoue / Shiho Amagasa (TBC) 15 minutes **Discussion of emerging learnings on success and challenges**
16:30 – 17:00	ACTIVE BREAK WITH A HEALTHY SNACK		
17:00–18:30 01:00 Japan (26/11) 16:00 UK 9.00 Arizona 8.00 California	**SESSION 4** **Learning from country experiences – *continued***	Session moderator Speakers from countries	**Country presentations *(continued):*** **5 Norway – Kan 3 (2020–2022)** Presenter: Jostein Steene- Johannessen 15 minutes **6 Finland – FinFit-population study 2021–2022** Presenter: Tommi Vasankari 15 minutes **7 Sweden – Health on Equal Terms: 2018 sub-study with accelerometers** Presenter: Marita Friberg 15 minutes **Group discussion of learnings and experiences**
18:30–18:45	**Close Day 1**	MOH / RIVM WHO	**Summary reflections** by Dutch MOH **Summary reflections** by WHO
18:45–20:15	**Networking dinner at Hotel Figi**		

Day 2: Wednesday 26 November 2025

Time	Item	Session moderator	Location: RIVM G22
07:30–08:15	BREAKFAST		
08.15–9:00	Transport to RIVM (*cycling 7 km*) & coffee/tea **NOTE: ID is needed to enter RIVM-premises**		
09.00–9.15	**SESSION 5**	WHO	Recap Day 1 and set context for Day 2
09:15–10:45 20.15 Fiji 17:15 Japan 8:15 UK 1.45 Arizona 0.45 California	**SESSION 6** **Learning from pilot studies**	Session moderator Speakers from countries	**Countries exploring use of wearables: pilot and feasibility studies** **8 – the Kingdom of the Netherlands** **Learnings from pilot studies** Presenter: Barbara Snoeker 20 minutes **9 – Active Lives Feasibility Study in England, UK** Presenter: Melvyn Hillson and Tessa Strain (video link) 15 minutes **Group discussion of learnings and experiences**
10:45–11:00	COFFEE / TEA break		
11:00–13:00 19:00 Japan 10:00 UK 3:00 Arizona 2:00 California	**SESSION 7** **Protocols and Methods**	**WHO**	**Discussion of "best practice" protocols and methods for use of wearable devices**
13:00–14:00	LUNCH & *time for a short walk*		

Time	Item	Session moderator	Location: RIVM G22
14:00–15:30 22:00 Japan 13:00 UK 6am Arizona 5am California	**SESSION 8** **Evidence on wearable device-based measurement of cycling**	Chairs: Annemarie Koster and WHO	Session Introduction **Presentation 1:** Results of rapid review of evidence on wrist work wearables and measurement of cycling **Presenter:** Dan Fuller **Presentation 2:** Evidence from improved machine learning algorithm for detecting cycling behaviours using two free-living validation datasets with younger (CAPTURE-24) and older (ELSA) British adults with wearable camera data **Presenter:** Laura Brocklebank **Presentation 3:** Results from harmonization of wrist and thigh accelerometer signals to improve cycling detection **Presenter:** Matthew Almadi **Presentation 4:** Results of independent testing – WAVES consortium data (title TBC) **Presenter**: Sarah Keadle **Group discussion** Does the current and emerging evidence support use of a wrist-worn wearable device with sufficient accuracy to detect cycling behaviours in minutes of a MVPA metric?
15:30–15:45	BREAK		
15:45–17:30 23:45 Japan 14:45 UK 7:45 Arizona 6:45 California	**SESSION 9** **Wearable device-based measurement of cycling behaviours**	Chairs: Annemarie Koster and WHO Daniel/ Sarah Matthew	**Focus on cycling continued** **Presentation continues (as needed)** Questions to discuss: 1. Can devices "adequately" detect MVPA from cycling/differentiate cycling behaviours from other movements? 2. Open-source machine learning algorithm available which detects cycling? 3. Do we know enough? What are the knowledge gaps? 4. What are the priority research questions to support WHO drafting guidance on the use of wearables?
17:30–17:45	**Close Day 2**	RIVM / WHO	Summary reflections on Day 2
17:45–18:30	**Return to Hotel –** *cycling 7 km if weather permits*		
19:00	**Networking dinner at restaurant First – 2e Dorpsstraat 67 Zeist –** 2 min walk from Hotel Figi		

Day 3: Thursday 27 November 2025

Time	Item	Session moderator	Location: FIGI Cinema
09.00–09.15	**Session 10**	RIVM	**Opening Day 3 & welcome** by Director General Sicco Louw
09.15–10.45 17:15 Japan 08:15 UK 01.15 Arizona 00:15 California	**Session 11** Global research priorities and opportunities	WHO	**WHO – Introduction to Day 3** **Presentation 1:** Review of research gaps, priorities and opportunities (see Montreal meeting report) **Presentation 2: WAVES consortium** – Study 1 – Independent evaluation of current open-source algorithms (Dan Fuller) **Q&A and discussion – including other research opportunities and collaborations**
10:45–11:00	COFFEE / TEA break		
11:00–12:15 19:00 Japan 10:00 UK 3.00 Arizona 2:00 California	**Session 12**		**Placeholder – space for carry forward issues** *Potential use* *i. for further discussion on surveillance systems protocols and methods* *ii. further review of device specifications and data analytics protocols (see Appendix tables Montreal Report)*
12:15–13:00	LUNCH		
13:00–14:00 21:00 Japan 12:00 UK 5.00 Arizona 4:00 California	**SESSION 13** **Data sharing platforms**	Chair: RIVM	**Presentation: Overview of Sports Data Valley** (https://info.sportdatavalley.nl/) – a platform for sharing monitoring data enabling secondary analyses in the Netherlands. Presenters: Remco Hoekman & Dennis van Kooij Mulier Institute **Q&A and discussion on useability**
14:00–15:15 22:00 Japan 13:00 UK 6:00 Arizona 5:00 California	**SESSION 14** **Closing session**	WHO and RIVM	**Recap of meeting** • Summary of group recommendations • Overview of next steps • Summary comments from delegates
15:15–15:30	**Closing**	MOH, RIVM & WHO	Thank you & farewell
15:30	**End of meeting** & transport to train station / airport		

Annex 2: List of participants

Government representatives

Canada

Dr Rachel Colley
Health Analysis Division
Statistics Canada
Canada

Dr Stephanie A. Prince Ware
Centre for Surveillance and Applied Research
Public Health Agency of Canada
Canada

Ms Karen Roberts *(virtual)*
Centre for Surveillance and Applied Research
Public Health Agency of Canada
Canada

Ms Pamela Turpin *(virtual)*
Centre for Chronic Disease Prevention and Healthy Equity
Public Health Agency of Canada
Canada

Denmark

Mr Tue Kristensen
The Danish Health and Medicines Authority
Denmark

Dr Mette Toftager
Research Unit of Active Living
Department of Sports Science and Clinical Biomechanics
University of Southern Denmark
Denmark

Finland

Dr Jouni Lahti
Department of Public Health and Welfare
Finnish Institute for Health and Welfare
Finland

Dr Tommi Vasankari
UKK Institute for Health Promotion Research
Faculty of Medicine and Health Technology
Tampere University
Finland

Japan

Dr Shiho Amagasa
Graduate School of Public Health
Teikyo University
Japan

Dr Shigeru Inoue
Department of Preventive Medicine and Public Health
Tokyo Medical University
Japan

Dr Rei Ono *(virtual)*
National Institutes of Biomedical Innovation, Health, and Nutrition (NIBN)
Japan

Kingdom of the Netherlands

Dr Maaike Kompier
CBS Statistics Netherlands
Netherlands

Mr Bram ter Schegget
Department of Sports and Physical Activity
Ministry of Health, Welfare and Sport
Kingdom of the Netherlands

Norway

Mr Olov Belander
Norwegian Directorate of Health
Norway

Dr Ulf Ekelund
Norwegian Public Health Institute
Department of Sports Medicine
Norwegian School of Sport Sciences
Norway

Dr Jostein Steene-Johannessen
Department of Sports Medicine
Norwegian School of Sport Sciences
Norway

Sweden

Dr Marita Friberg
Department of Living Conditions and Lifestyles
Public Health Agency of Sweden
Sweden

Dr Maria Hagströmer
Division of Physiotherapy
Karolinska Institutet
Sweden

United Kingdom of Great Britain and Northern Ireland

Dr Melvyn Hillsdon *(virtual)*
NIHR Exeter Biomedical Research Centre
Exeter University
United Kingdom of Great Britain and Northern Ireland

Ms Helen Price *(virtual)*
Sport England
United Kingdom of Great Britain and Northern Ireland

Dr Tessa Strain *(virtual)*
Moray House of Education and Sport
Institute for Sport, PE and Health Sciences
University of Edinburgh
United Kingdom of Great Britain and Northern Ireland

External academic experts

Dr Matthew Ahmadi *(virtual)*
Physical Activity, Lifestyle, and Population Health
Mackenzie Wearables Research Hub
Faculty of Medicine and Health
The University of Sydney
Australia

Dr Kerstin Bach *(virtual)*
Data and Artificial Intelligence Group
Department of Computer Science
Norwegian University of Science and Technology
Norway

Dr Laura Brocklebank
Nuffield Department of Population Health
University of Oxford
United Kingdom of Great Britain and Northern Ireland

Dr Aiden Doherty
Nuffield Dept. of Population Health
University of Oxford
United Kingdom of Great Britain and Northern Ireland

Dr Thijs Eijsvogels
Exercise Physiology Research Group
Department of Medical BioSciences
Radboud University Nijmegen
Kingdom of the Netherlands

Dr Daniel Fuller
Community Health and Epidemiology
College of Medicine
University of Saskatchewan
Canada

Dr Marco Giurgiu
Department of Sport and Sport Science
Karlsruhe Institute of Technology
Germany

Dr Remco Hoekman
Faculty of Social Sciences
Radboud University Nijmegen
Mulier Institute
Kingdom of the Netherlands

Dr Sarah Kozey Keadle *(virtual)*
Department of Kinesiology and Public Health
California Polytechnic State University
United States of America

Mr Dennis van Kooij
Mulier Institute
Kingdom of the Netherlands

Dr Annemarie Koster
Department of Social Medicine
Maastricht University
Kingdom of the Netherlands

Dr Ben Maylor
Nuffield Department of Population Health
University of Oxford
United Kingdom of Great Britain and Northern Ireland

Dr Hidde van der Ploeg
Department of Public and Occupational Health
Amsterdam UMC, VU University Amsterdam
Kingdom of the Netherlands

Dr Manos Stamatakis
Physical Activity, Lifestyle, and Population Health
Mackenzie Wearables Research Hub
Faculty of Medicine and Health,
School of Health Sciences
The University of Sydney
Australia

Dr Rick Troiano
Retired (National Institutes of Health)
Surveillance of Physical Activity Expert
Maryland
United States of America

Ms Inge de Wolf
Department of Public and Occupational Health of Amsterdam UMC
University of Amsterdam
Kingdom of the Netherlands

National Institute for Public Health and the Environment (RIVM)

Dr Barbara Snoeker
Centre for Prevention, Lifestyle and Health
National Institute for Public Health and the Environment
Kingdom of the Netherlands

Ms Annemarie van der Vegt
Centre for Prevention, Lifestyle and Health
National Institute for Public Health and the Environment
Kingdom of the Netherlands

Dr Wanda Wendel-Vos
Centre for Prevention, Lifestyle and Health
National Institute for Public Health and the Environment
Kingdom of the Netherlands

World Health Organization

Dr Fiona Bull
Physical Activity Promotion and Alcohol Harm Prevention Unit
Department of Health Determinants, Promotion and Prevention
World Health Organization
Switzerland

Dr Leanne Riley
Surveillance, Monitoring and Reporting Noncommunicable Diseases Unit
Department of Noncommunicable Diseases & Mental Health
World Health Organization
Switzerland

Dr Ivo Rakovac
Noncommunicable Diseases Surveillance
Regional Office for Europe
World Health Organization
Denmark

Dr Mariken Leurs
Department of Health Determinants Promotion and Prevention
World Health Organization
Switzerland

Dr Tomo Kanda *(virtual)*
Pacific Island Countries
Regional Office for the Western Pacific
World Health Organization
Fiji

Dr Roberta de Betania Caixeta *(virtual)*
Noncommunicable Diseases Surveillance, Prevention and Control
Regional Office for the Americas
World Health Organization
United States of America

Dr Heba Fouad *(virtual)*
Noncommunicable Diseases Surveillance
Regional Office for the Eastern Mediterranean
World Health Organization
Egypt

All participants completed and submitted a WHO declaration of interest form. Unless listed below, delegates reported no conflicts.

- Dr Aiden, Dr Brocklebank and Dr Maylor – reported that their research group receives support the following entities with an interest in the subject of the meeting: Novo Nordisk, Boehringer, Ingelheim, SwissRe, Google and GSK.
- Dr Fuller – reported having received grant funding by the Canadian Institute of Health research on the topic and consulting on research projects using commercial wearable devices for the National Institute of Health, United States of America.

- Dr Hillson – reported having proprietary knowledge and that his employer (the University of Exeter) has a collaborative agreement with Activinsights, the manufacturer of the GENEActive accelerometer.
- Dr Keadle – reported active grant funding through the National Institute of Health, United States of America, and contracting work for Sentimetrix. The company filed for a provisional patent.
- Dr Stamatakis – shareholder of a company based in the United States of America (business-to-business services enterprise) and paid expert witness in current trial.

None was deemed to represent a conflict of interest.

Annex 3: Summary country templates

Full summary of methods and protocols for use of wearable device in adult populations in national health monitoring systems is available on request from letsbeactive@who.int

Table A3.1. Country system/study overview

Question[a]	1.1 System/study name	3.1 Device	3.4 Wear location	1.6 Recent year	1.5 First year	1.7 No. of times	1.4 Partners	3.3 Device procurement	4.2 (sub)sample (initial[c])
Canada	Canadian Health Measures Survey (CHMS)	Actigraph[b] wGT3x-BT	Waist/hip	2025–27	2007	8 cycles	None	purchased	3264
Denmark 1	Danish National Health Survey (DNHS)	SENSmotion	Thigh	2023	2023	1	Express/AS Sensmotion University	Purchased	6993
Denmark 2	Moving Denmark 2020	Axivity AX3	Thigh	2021/22	2021/22	1	University	Purchased	1248
Finland 1	Healthy Finland 2022–23	Actigraph GT9X Link	Wrist, ND*	2023	2017	2	University	Purchased	1980
Finland 2	FINFIT STUDY	UKK- RM42	Waist, ND*+ Wrist at night	2025–26	2017–18	3	None	Own device	10 500
Japan	National Health and Nutrition Survey (NHNS)	Yamax (pedometer)	Waist/hip (front pocket from 2024)	2024	1989	34	Yamasa Corp NIBN	Purchased	4272
Norway	Norwegian National Physical Activity Surveillance (NNPAS)	Actigraph GTX3X+BT	Waist/hip	2022/22	2008/09	3	Ministry & University & NIPH	Purchased	6089
Sweden	Health on Equal Terms (HLV)	UKK – RM42	Waist, ND*+ Wrist at night	2025	2018	2	Ministry & UKK + others	Borrowed	4000

Question[a]	1.1 System/study name	3.1 Device	3.4 Wear location	1.6 Recent year	1.5 First year	1.7 No. of times	1.4 Partners	3.3 Device procurement	4.2 (sub)sample (initial[c])
United States of America	National Health & Nutrition Examination Survey (NHANES)	Actigraph GT3x	Wrist – ND*	2011/14	2003/06	2	NIH, CDC	Purchased	16 000
Kingdom of the Netherlands	Accelerometer pilot	ActivPAL 3/4	Thigh	2024	2023	pilot	RIVM, CBS ++	Purchased	1100

[a] Refers to the corresponding question number in the template – see Web annex; [b] Actical device cycles 1–6, Actigraph cycle 7 from 2022; [c] Rough indication of initial/starting device sample;
***** ND non-dominant wrist

Table A3.2. Administration

	1.1 System/study name	3.1 Device	3.6/3.7 Device Feedback?	2.9 Handouts?	2.10 Incentive?	5.1 Device handout	5.2 Device return	3.9 Use of ques/diary (yes or no)	3.10 Time overlap[a]
Canada	Canadian Health Measures Survey (CHMS)	Actigraph wGT3x-BT	No screen No feedback	Yes	Yes – $	In person centre	Post return	Q-Y D-N	No
Denmark 1	Danish National Health Survey (DNHS)	SENSmotion	No screen No feedback	Yes	Yes – raffle	Posted to home	Post return	Q-Y D-Y	Yes
Denmark 2	Moving Denmark 2020	Axivity AX3	No screen No feedback	Yes	Yes – own data	Posted to home	Post return	Q-Y	Yes
Finland 1	Healthy Finland 2022–23	Actigraph GT9X Link	No feedback + time of day	Yes	Yes – own data	Posted to home	Post return	Q-Y D-Y	Yes
Finland 2	FINFIT STUDY	UKK-RM42	Screen – Yes + time of day	Yes	No	In person center	Post return	Q-Y D-Y	No – Ques Yes – Sleep diary
Japan	National Health and Nutrition Survey (NHNS)	Yamax (pedometer)	Screen – Yes Feedback – steps	Yes	Yes – the pedometer	In person home	Not applicable	Q-Y D-N	Yes
Norway	Norwegian National Physical Activity Surveillance (NNPAS)	Actigraph GTX3X+BT	No screen No feedback	Yes	Yes – $ voucher	Posted to home	Post return	Q-Y D-Y	Yes
Sweden	Health on Equal Terms	UKK-RM42	No screen No feedback	Yes	Yes – own data	Posted to home	Post return	Q-Y D-Y	Yes
United States of America	National Health & Nutrition Examination Survey (NHANES)	Actigraph GT3x	No screen No feedback	Yes	Yes – $	In person centre	Post return	Q-Y	No
Kingdom of the Netherlands	Accelerometer pilot	ActivPAL 3/4	No screen No feedback	Yes	Yes – $ voucher	Posted to home	Post return	Q–Y D-N	No

[a] The time period of the questionnaire/diary corresponded with the wear time of device.

Table A3.3. Requirements and training (and indicative sample size)

	1.1 System/study name	3.1 Device	2.8 Wear time instructions	6.11 Recharge required	3.5 Other sensors?	5.5. Device problems %	6.2 Data download	8.1 Admin/ data training?
Canada	Canadian Health Measures Survey (CHMS)	Actigraph wGT3x-BT	7/24	No	Gyroscope PPG	Malfunction: 2%	StatCan Secure Server	Admin-Y Data-N
Denmark 1	Danish National Health Survey (DNHS)	SENSmotion	7/24	No	Gyroscope	Very few	Amazon webserver	Admin-Y Data-N
Denmark 2	Moving Denmark 2020	Axivity AX3	7/24 – RW[a]	No	Skin sensor	Minor – few	University server	Admin-Y Data-N
Finland 1	Healthy Finland 2022–23	Actigraph GT9X Link	7/24 – RW[a]	No	No	Battery	Secure server in THL	Admin-N Data-N
Finland 2	FINFIT STUDY	UKK – RM42	7/24 – RW[a] + switch[b]	No	Y – posture metrics	None	UKK server	Admin-Y Data-N
Japan	National Health and Nutrition Survey (NHNS)	Yamax (pedometer)	1 day waking hours	No	No	No info	Manual diary data entry	Admin-Y Data-N
Norway	Norwegian National Physical Activity Surveillance (NNPAS)	Actigraph GTX3X+BT	7/waking hours only – RW[a]	No	No	2–3%	University server	Admin-Y Data-N
Sweden	Health on Equal Terms	UKK – RM42	7/24 -RW[a] + switch[b]	No	No	Few	University & PH agency	Admin-N Data-N
United States of America	National Health & Nutrition Examination Survey (NHANES)	Actigraph GT3x	7/24	No	Light sensor	Some detected on return	Contractor & NCHS	Admin-Y Data-Y
Kingdom of the Netherlands	Accelerometer pilot	ActivPAL 3/4	9 days / 24 RW[a]	No	Gyroscope PPG	Some battery issues	At survey center & RIVM	Admin-Y Data-N

[a] RW – Instructed to remove for water-based activity/shower/bath; [b] Switch = instruction for different wear location at night.

Table A3.4. Data handling and analysis

	1.1 System / study name	3.1 Device	*6.3* Raw data?	6.5 Algorithm	6.4 Minimum wear time for analysis	6.7 Nr/bias Adjustment	6.8 Sub-pop analysis?
Canada	Canadian Health Measures Survey (CHMS)	Actigraph wGT3x-BT	*Yes*	Developed own in SAS (Accel) and in R (Actical)	≥ 4 days & ≥ 10 hrs	Yes	Yes
Denmark 1	Danish National Health Survey (DNHS)	SENSmotion	Yes-P*	P* Actimotus	3 days & ≥ 20 hours + 5 min walk/day	No	No
Denmark 2	Moving Denmark 2020	Axivity AX3	Yes	Developed own	≥ 4 days & ≥ 10 hours + 1WEND[a]	Yes	No
Finland 1	Healthy Finland 2022–23	Actigraph GT9X Link	Yes	GGIR R-package, Hildebrand / van Hees algorithms	3 days – 24 hrs	Yes	Yes
Finland 2	FINFIT STUDY	UKK- RM42	Yes	Own – UKK MAD-APE	≥ 4 days & ≥ 10 hours* (+24h)[a]	Yes	Yes
Japan	National Health and Nutrition Survey (NHNS)	Yamax (pedometer)	No	N/A	< 100 or ≥ 50 000 steps on the day were excluded	No	Yes
Norway	Norwegian National Physical Activity Surveillance (NNPAS)	Actigraph GTX3X+BT	Yes	P* – Kinesoft	≥ 2 days & ≥ 10 hours	Yes	Yes
Sweden	Health on Equal Terms	UKK –RM42	Yes	Open source: MAD-APE	≥ 4 days & ≥ 10 hours + 1WEND[b]	No	No
United States of America	National Health & Nutrition Examination Survey (NHANES)	Actigraph GT3x	Yes	Own – MIMS	Various used – incL ≥ 1 day	Yes	No
Kingdom of the Netherlands	Accelerometer pilot	ActivPAL 3/4	Pre-processed	P* PALtechnologies	≥ 4 days / ≥ 20 hours & ≥ 500 steps	No	No

[a] Also required 24 hours for 24-hour compositional analysis; [b] Including one weekend day (WKEND), P*: Proprietary software

Table A3.5. Output metrics and reporting

	1.1 System / study name	3.1 Device	6.9 Step count	6.9 MVPA	6.9 Meet recommend.	6.9 Sedentary metric	6.9 Other?	6.11 Report both self-report & device
Canada	Canadian Health Measures Survey (CHMS)	Actigraph wGT3x-BT	Yes	Yes	Yes	Yes – ≤ 9 hours/ day	Yes – sleep (under devel)	Yes
Denmark 1	Danish National Health Survey (DNHS)	SENSmotion	Yes	Yes	Yes	Yes – mins sit/ stand/lie	yes – bike, sleep	Yes
Denmark 2	Moving Denmark 2020	Axivity AX3	Yes	Yes	Yes	Yes – mins sit/ stand/lie	Yes – run / walk / bike	Yes
Finland 1	Healthy Finland 2022–23	Actigraph GT9X Link	No	Yes	No	Yes – mins	Yes sleep and light PA	No
Finland 2	FINFIT STUDY	UKK- RM42	Yes	Yes	Yes	Yes – lying/ reclining/ sitting/standing	Yes – many	Yes
Japan	National Health and Nutrition Survey (NHNS)	Yamex (pedometer)	Yes	No	No	No	No	Yes
Norway	Norwegian National Physical Activity Surveillance (NNPAS)	Actigraph GTX3X+BT	Yes	Yes	Yes	Y – mins/day	No	Yes
Sweden	Health on Equal Terms	UKK – RM42	Yes	Yes	No	Y – mins/day[a]	No	Yes
United States of America	National Health & Nutrition Examination Survey (NHANES)	Actigraph GT3x	No	No	No	No	Ref curves	Yes
Kingdom of the Netherlands	Accelerometer pilot	ActivPAL 3/4	Yes	Yes	Yes	Yes total, naps, sitting	**Cycling mins, standing mins, sleep**	Yes

[a] Also reported % with 50% and 75% of waking time sedentary.

For further information, contact:
Department of Health Determinants, Promotion, and Prevention
World Health Organization
20 avenue Appia
1211 Geneva 27
Switzerland
Website: https://www.who.int/