



Measuring physical activity in adults using wearable technologies: **report of a scientific meeting**

10–12 June 2025
Montreal, Canada

Measuring physical activity in adults using wearable technologies: report of a scientific meeting

10–12 June 2025
Montreal, Canada

Measuring physical activity in adults using wearable technologies: report of a scientific meeting, 10–12 June 2025, Montreal, Canada

ISBN 978-92-4-011669-6 (electronic version)

ISBN 978-92-4-011670-2 (print version)

© World Health Organization 2025

Some rights reserved. This work is available under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 IGO licence (CC BY-NC-SA 3.0 IGO; <https://creativecommons.org/licenses/by-nc-sa/3.0/igo>).

Under the terms of this licence, you may copy, redistribute and adapt the work for non-commercial purposes, provided the work is appropriately cited, as indicated below. In any use of this work, there should be no suggestion that WHO endorses any specific organization, products or services. The use of the WHO logo is not permitted. If you adapt the work, then you must license your work under the same or equivalent Creative Commons licence. If you create a translation of this work, you should add the following disclaimer along with the suggested citation: “This translation was not created by the World Health Organization (WHO). WHO is not responsible for the content or accuracy of this translation. The original English edition shall be the binding and authentic edition”.

Any mediation relating to disputes arising under the licence shall be conducted in accordance with the mediation rules of the World Intellectual Property Organization (<http://www.wipo.int/amc/en/mediation/rules/>).

Suggested citation. Measuring physical activity in adults using wearable technologies: report of a scientific meeting, 10–12 June 2025, Montreal, Canada. Geneva: World Health Organization; 2025. Licence: [CC BY-NC-SA 3.0 IGO](https://creativecommons.org/licenses/by-nc-sa/3.0/igo).

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

Sales, rights and licensing. To purchase WHO publications, see <https://www.who.int/publications/book-orders>. To submit requests for commercial use and queries on rights and licensing, see <https://www.who.int/copyright>.

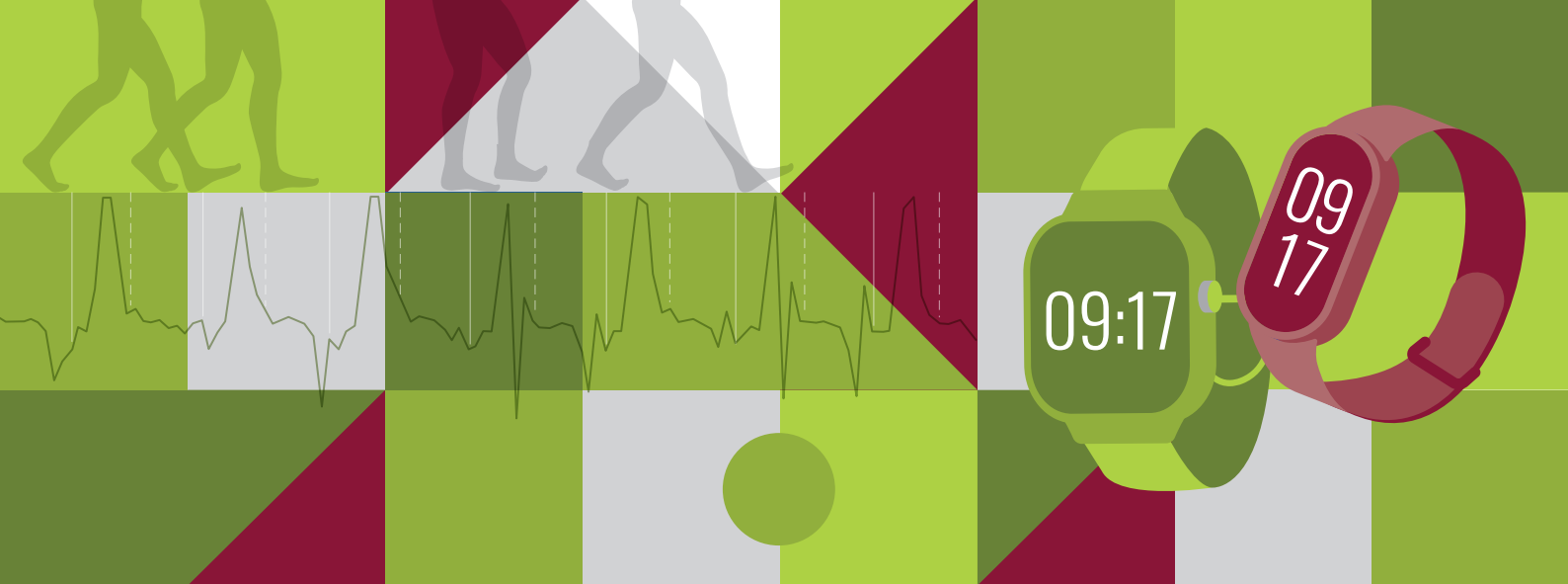
Third-party materials. If you wish to reuse material from this work that is attributed to a third party, such as tables, figures or images, it is your responsibility to determine whether permission is needed for that reuse and to obtain permission from the copyright holder. The risk of claims resulting from infringement of any third-party-owned component in the work rests solely with the user.

General disclaimers. The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of WHO concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement.

The mention of specific companies or of certain manufacturers' products does not imply that they are endorsed or recommended by WHO in preference to others of a similar nature that are not mentioned. Errors and omissions excepted, the names of proprietary products are distinguished by initial capital letters.

All reasonable precautions have been taken by WHO to verify the information contained in this publication. However, the published material is being distributed without warranty of any kind, either expressed or implied. The responsibility for the interpretation and use of the material lies with the reader. In no event shall WHO be liable for damages arising from its use.

Design and layout by Inis Communication



Contents

Acknowledgements	vi
Executive summary	1
Background	5
Objectives of the meeting	7
Day 1	8
Session 1: Welcome, land acknowledgement, and overview (Dr Karim Khan, Dr Fiona Bull)	8
Session 2: What are we measuring? (Dr Fiona Bull, Dr Stephanie Prince Ware, Dr Rachel Colley)	10
Session 3: What is this device going to do? What must it have, and what is nice to have? (Dr Fiona Bull, Dr Rick Troiano, Dr Matthew Ahmadi)	15
Session 4: Validating, storing, analysing, and harmonizing data (Dr Aiden Doherty, Dr Manos Stamatakis, Dr Kerstin Bach, Dr Marco Giurgiu, Dr Sarah Keadle, Dr Charles Matthews)	19
Day 2	29
Session 5: Follow-on work on specifications of device attributes, protocols and data analysis – continued with report-back from sub-group (Dr Aiden Doherty)	29
Session 6: Advancing evaluation evidence (Dr Sarah Keadle and Dr Fiona Bull)	32

Session 7: Age-related considerations and sedentary behaviours (Dr Marla Beauchamp)	35
Session 8: What are the likely next steps following this meeting? (Dr Fiona Bull)	37
Closing Day 2: Commentary from an industry perspective: Discovery Vitality (Dr Mosima Mabunda)	39
Day 3	41
Session 9: A framework for developing consensus (Paul Blazey)	41
Session 10: Report on ongoing process towards developing guidelines for reporting research on wearable devices (Dr Marco Giurgiu, Dr Jason Gill)	42
Session 11: Closing plenary (Dr Fiona Bull)	45
Closing	46
References	47
Annex 1: Provisional programme	49
Annex 2: List of participants	52
Annex 3: Discussion on factors that influence comparability of data on physical activity and sedentary behaviours from wearable devices	56
Annex 4: Reasons for the transition from self-report to wearables	71

Acknowledgements

This report presents outputs from the third in a series of meetings convened in support of WHO's work to develop new guidance on the use of wearable devices to measure physical activity in national population health surveillance systems.

The meeting, held in Montreal in June 2025, was jointly conducted by the Physical Activity Unit, Department of Health Promotion at WHO Headquarters; and the Canadian Institutes of Health Research (CIHR) Institute of Musculoskeletal Health and Arthritis (IMHA).

The meeting report was prepared by Dr Fiona Bull, Head of the Physical Activity Unit, WHO with contributions from Dr Rick Troiano (WHO consultant) and input from CIHR-IMHA team members Dr Hetty Mulhall and Eunice Lui and all meeting delegates. The report was reviewed by all meeting delegates.

WHO extends its thanks to the delegates who attended this meeting for their contributions during the meeting and in the compilation of this report. WHO also extend deep thanks to the leadership of Dr Karim Khan and CIHR-IMHA for providing financial support to host and coordinate the logistics of the meeting. Additional thanks are extended to Professor Manos Stamatakis who provided additional funding to support the attendance of some delegates.

Executive summary

Background

The third WHO Technical Meeting on Physical Activity Measurement and Surveillance (Montreal, 10–12 June 2025) brought together global experts to advance the development of guidance on integrating wearable devices into national population health surveillance systems. Over three days, participants focused on defining core metrics, device specifications, and protocols for use; assessing the evidence base for evaluation of wearable devices; and discussing knowledge gaps and research needs that need addressing to improve measurement and tracking of trends in physical activity and sedentary behaviours.

The background for the meeting is the global public health challenge of increasing levels of physical activity. Regular physical activity is essential for physical and mental health, yet over 80% of adolescents and 31% of adults fail to meet WHO recommended levels. To monitor progress towards the 15% relative reduction of physical inactivity target by 2030, as set out in the Global Action Plan on Physical Activity (GAPPA), regular, robust and reliable population-level surveillance is essential.

Today, most countries rely on self-report surveys, which have well-documented limitations. By contrast, wearable devices such as wrist- or hip-worn accelerometers offer more accurate and continuous data on physical activity and sedentary behaviour (and potentially sleep, too). However, methods and instruments vary widely, limiting comparability of data.

No global guidance exists on appropriate devices, protocols, or analytic approaches, and technical, financial, and logistical barriers – especially in low- and middle-income countries – constrain adoption. Developing global guidance on the use of wearable devices in health monitoring systems is therefore a priority under WHO's 14th General Programme of Work.

Two previous meetings scoped the current contexts on use of wearable devices in adults (2023) (1) and adolescents (2024) (2). The Montreal discussions focused on developing device specifications for use among adults and older adults, drawing on expertise in epidemiology, wearable technologies, algorithm development, and national surveillance systems.

Meeting objectives

Participants were invited to take part in the following activities.

- Agree on a core set of output metrics from wearable devices for adult and older adult populations.
- Review and contribute to a draft working document on wearable device specifications that are “fit for purpose” for national health monitoring systems tracking adults and older adult populations.
- Assess the status and needs for open-source algorithms, validation studies, and harmonization of data collection and analytic approaches.
- Discuss complementary data needs, integration with survey instruments, and considerations for older populations.
- Identify evidence gaps and priority research to inform drafting of WHO guidance and country implementation.

Key discussions and outcomes

1. **Core metrics:** Delegates agreed three feasible and appropriate metrics for near-term guidance: (i) moderate-to-vigorous physical activity (MVPA); (ii) step count; and (iii) sedentary behaviour. These align with WHO’s 2020 Guidelines on physical activity and sedentary behaviour (3), and/ or anticipate future expansion. Sleep and other specific physical activities (e.g. muscle strengthening) were acknowledged as important but out of scope.
2. **Device specifications:** Consensus was reached on many essential (minimum) features for a wearable device for use in national population surveillance systems: non-dominant wrist placement; minimum seven days of wear; triaxial accelerometry; and mandatory access to raw data. Other parameters such as wear time, sampling frequency, epoch length, and the inclusion of additional sensors were discussed within the context of balancing rigour and precision with affordability and feasibility, especially for low- and middle-income countries. The “good enough” principle was applied to develop a set of recommendations to WHO.
3. **Analytic methods and algorithms:** Evidence from the UK Biobank study, ProPASS, and the HUNT study show that open-source methods can produce robust estimates, particularly for steps and sedentary time. It was agreed that analysis of MVPA remains more challenging and further research is required to improve algorithms and validation, especially in older adults. Delegates strongly supported the need for independent evaluation studies, harmonization of methods, and development of pooled training datasets.

4. **Validation evidence:** It was acknowledged that existing studies are uneven in quality, sample size, and diversity and are mostly limited to high-income contexts. More research is needed across the range of socioeconomic, ethnic, and functional (mobility) populations, including low- and middle-income countries and older adults. Independent evaluation of methods, standardized definitions, and demonstration in low- and middle-income countries were identified as essential for building global uptake.
5. **Country experiences:** Canada's experiences with the Health Measures Survey illustrated how wearable devices can be successfully integrated into national surveillance system, and lessons were shared on protocol transitions, cross-over validation, and participant compliance. Such experiences demonstrate both potential and logistical challenges and provide models for other countries. It was recommended that further meetings should be convened to learn from other country experiences should be undertaken.
6. **Older adults:** Special considerations are required for populations with altered gait, mobility aids, or neurodegenerative conditions. Algorithms trained on younger adults may misclassify behaviours. Inclusive terminology (e.g. "step or step-equivalents") and population-specific validation were recommended.
7. **Next steps:** WHO outlined a roadmap including: (i) finalizing the report of the Montreal meeting; (ii) drafting and consulting on Version 1 of WHO's technical guidance; (iii) establishing mechanisms to support a global research consortium; (iv) resource mobilisation to support conducting independent evaluations of existing algorithms; and (v) planning multi-country studies, particularly in LMICs, to strengthen evidence and support phased adoption.

The following research priorities were identified:

- a. Independent evaluation of existing algorithms on MVPA, sedentary behaviour, and step-count measurement.
- b. More developmental research to improve both machine learning methods and other methods to classify and develop an MVPA metric and to evaluate performance in older adults and more diverse and low- and middle-income populations.
- c. Evaluation studies in multiple low- and middle-income countries to test and refine algorithms in populations.
- d. Demonstration studies to field test the utility and feasibility of WHO draft guidance on the use of wearable devices in population health surveillance systems across set of diverse cultures and contexts.

Conclusion

The Montreal meeting confirmed both the scientific readiness and policy urgency for WHO to lead development of global guidance on wearable devices for physical activity surveillance. While evidence gaps remain – particularly for MVPA, older adults, and in low- and middle-income country contexts – currently available analytic methods are sufficient to begin adoption and a transition from or adjunct to self-report surveys. Overall delegates agreed that with pragmatic global guidance, additional collaborative research, and conducting country demonstration projects, WHO and its partners could enable countries to adopt wearable device-based surveillance, strengthen national health data information systems, and accelerate policy implementation to make progress towards the 2030 global target on physical inactivity.

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 WHO's recommended levels of physical activity (4).

To address this challenge, the Global Action Plan on Physical Activity 2018–2030 (5) (GAPPA) provides countries with a set of evidence-based policy recommendations. When implemented through a coordinated, multisectoral approach, these actions can help increase physical activity levels and contribute to broader 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 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. These include recall inaccuracy, response bias, and high respondent burden (for example because of complexity of questions and length). Wearable digital devices offer an alternative with several advantages: they enable continuous, multi-day data collection across different movement types; are non-invasive; and place minimal burden on respondents. Therefore, wearable device-based measurement offers a method providing more robust and more comprehensive data. However, technical, logistical, and financial barriers currently limit their adoption and routine use at scale in national health surveillance systems.

Importantly, there is no internationally agreed standard – nor WHO-endorsed guidance – on which wearable devices are appropriate, or which data collection and analysis protocols should be used to report global physical activity metrics. Moreover, the higher cost of advanced wearable devices presents a particular challenge for large-scale adoption in low- and middle-income countries. Nonetheless, there is growing interest in identifying affordable, practical, wearable technologies that could support national monitoring of physical activity, sedentary behaviour, and sleep.

Strengthening health information systems and harnessing digital technologies are recognized priorities for global health. In this context, the development of WHO global guidance on the use of wearable devices for physical activity surveillance is a priority workstream for WHO's Physical Activity Unit and a formally approved global technical product for the 2024–2025 biennium. This work is also aligned with advancing WHO's 14th General Programme of Work (GPW14) – the organization's strategic framework for guiding its work over four years.

This report summarizes the third in a series of expert technical meetings convened by WHO. This meeting (held 10–12 June 2025) brought together global experts in measurement science, wearable technologies, and public health surveillance. It complements two prior meetings: a scoping consultation in November 2023 focused on adults (1), and a meeting in May 2024 (2) that focused specific attention on adolescent populations.

This third meeting aimed to advance these efforts. Specifically, it aimed to draw on scientific evidence and expert experience to inform the development of global guidance on the use of wearable devices in the monitoring and surveillance of physical activity and sedentary behaviour across adult and older adult populations.

Objectives of the meeting

1. Gain feedback on the draft outline of a global guidance tool for the use of wearable devices in population health monitoring systems to assess physical activity and sedentary behaviour in adult, and older adult populations.
2. Agree on a set of primary output metrics for wearable devices used in population-based monitoring of physical activity and sedentary behaviour; and develop draft specifications of the core features and functions of a “fit for purpose” wearable device, suitable for use across different age groups, to measure physical activity and sedentary behaviours.
3. Review and discuss: (i) availability and accuracy of open-source algorithms; (ii) the need for and process of developing a globally agreed open-source method for processing raw data from wearable devices; and (iii) the validation requirements and approaches.
4. Discuss further data needed in combination with data obtained from a wearable device; and to identify how such data could be collected using supporting survey instruments or alternative methods.
5. Share and review current efforts to strengthen the scientific reporting of research and development on wearable technologies in the fields of physical activity and sedentary behaviours, and progress towards the development of standards for reporting.
6. Discuss and identify any unique differences and implications in the use of wearable devices to monitor physical activity and sedentary behaviour in older adult populations; and share developments from other initiatives seeking to develop consensus on methods.

Day 1

Session 1: Welcome, land acknowledgement, and overview (Dr Karim Khan, Dr Fiona Bull)

After a welcome including a traditional acknowledgement of the indigenous owners of the land on which the meeting was taking place, participant declarations of interest,¹ introductions, and sharing of expected meaningful outcomes from the meeting, Dr Bull introduced WHO, the background of the meeting, and outlined WHO's programme of work to update global guidance on the measurement and surveillance of physical activity across all age groups. Key points included:

- WHO is the UN's specialized agency for health, with 194 Member States (which are countries) and the goal of supporting the attainment by all peoples of the highest possible level of health.
- WHO core functions most relevant to this meeting are: setting global norms and standards such as guidelines; supporting global monitoring and assessing health trends; and shaping the research agenda.
- Physical activity is now embedded in global policy on NCD prevention and control – driven by scientific evidence and data and accomplished by the many roles and advocacy undertaken by researchers and civil society.
- GAPPA set out policy recommendations and the global target of reducing physical inactivity by a relative 15% by 2030.
- Recent data show that 31% of adults are not meeting global recommendations and that the world is “off track” to meet the target – most countries will not reduce physical inactivity by a relative 15% by 2030 (using 2010 as baseline).
- The focus of this meeting supports WHO's work to achieve GAPPA Strategic Objective 4, policy recommendation 4.2, which calls for *“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”*; and policy recommendation 4.3 that calls for *“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”*.

¹ Participant statements are provided in Annex 2.

WHO introduced the global NCD and risk factor surveillance system recommended for use in countries to monitor the health of adult populations known as STEPS – a stepwise approach.² Key details included:

- WHO recommends to countries that STEPS is used to collect data every 5 years.
 - Step 1 is usually a survey conducted at household level that includes the Global Physical Activity Questionnaire (GPAQ).
 - Step 2 involves collecting physical measures (often at household level) using standardized protocols and measurement.
 - Step 3 involves collecting biochemical measures (usually conducted in clinic-based locations).
- Optional STEPS modules on different topics are available to countries.
- STEPS already includes the option of adding a device-based measure of physical activity, using existing available guidance based on a WHO pilot study conducted in Malawi (updating these protocols and methods is the focus of WHO's Physical Activity Unit workplan).
- WHO's NCD surveillance technical team provides technical support to countries on request, across all phases of STEPS implementation – from planning, sampling and data collection to data analysis and reporting. When needed, WHO provides on-site training for country STEPS teams.
- Typical sample size is around 6000 adults aged 18–69 years (older ages are included if the country requests it) and data collection can take from around 6 weeks up to 3–4 months depending on country circumstances.
- WHO is usually working with 20–30 countries each year, all at different stages of STEPS implementation.

The most frequently reported difficulties in using the STEPS GPAQ include: the length of time it takes to complete; the questions seem similar and repetitive; the questions are confusing in some parts; respondents find it difficult to recall the duration and frequency of physical activity participation; and respondents also find it hard to differentiate between questions asking about levels of intensity. It was also noted that GPAQ has not been tested for use with people living with disabilities.

For national monitoring of physical activity in young people (specifically adolescents aged 11–17 years), WHO recommends the use of the Global School Health Questionnaire – further details were not provided as this age group was outside the scope of this meeting.

.....
² <https://www.who.int/teams/noncommunicable-diseases/surveillance/systems-tools/steps>

Purpose of this meeting

WHO reminded participants that the focus of the meeting was on adults and the measurement of physical activity and sedentary behaviour as part of national surveillance studies. Whilst sleep behaviour and measuring physical activity among children aged under 17 years as well as persons living with disabilities were also important, these groups will be addressed in detail in separate future meetings. It was also noted that the use of wearable devices in research versus population health surveillance systems may necessitate other considerations not discussed as part of this meeting.

WHO outlined that the primary objective was to provide WHO with expert input on the following four issues related to the use of wearable devices in population surveillance systems. A fifth issue (E) was noted but deemed beyond scope of the current meeting:

- A. What core metrics should wearable devices be used to report?
- B. What are the essential device specifications and recommended wear protocols?
- C. What (open-source) methods are currently available to analyse data from wearable devices?
- D. What, if any, other data should be collected alongside that provided by the wearable device?
- E. What methods and processes should a country use to transition from their current use of self-report to device-based estimates of physical activity in national reporting?

WHO noted that to advance the use of wearables in population surveillance, striving for perfection should not be a barrier to making progress. Wearable technology will continue to improve, but to commence building wider use outside of research contexts, WHO needs to assess if there is sufficient evidence that data collected from wearable devices are trustworthy and useful. And if so, is there sufficient knowledge to inform the development of a “good enough” protocol for wearable devices that will provide a clear improvement and an alternative to current practice using self-report questionnaires?

Session 2: What are we measuring? (Dr Fiona Bull, Dr Stephanie Prince Ware, Dr Rachel Colley)

This session focused on the first part of Meeting Objective 2. WHO reminded participants that the focus of this meeting was on population surveillance, which has recognized challenges such as: time constraints due to the broad nature of a health survey; logistical demands, including ensuring a representative sample and collecting data in rural areas and among equity groups; and the need for consistent, reliable measures over time to allow reporting of key metrics and trend assessment.

WHO noted that both GAPPA and the WHO Guidelines (2020) (3) provide clear direction on what measurement of physical activity in adults is required, namely MVPA across the

week; vigorous-intensity physical activity; and multicomponent physical activity³ for the prevention of falls; and muscle strengthening activities. In addition, both GAPP and the WHO Guidelines also call for measurement and reporting on sedentary behaviours.

WHO suggested that to make progress it was necessary to narrow the scope of discussion on methods for wearable devices to a small set of clearly defined output metrics that would be most feasible to implement in the near term. The potential to add further metrics in the future was noted due to opportunities to undertake retrospective analyses of raw data collected.

WHO advised that it was also important to look forward and anticipate emerging physical activity and sedentary behaviour measurement and surveillance needs, as new guidelines will be developed by WHO for publication in 2030. However, this should not distract from focusing on how to advance the use of wearables to track and report against current global guidelines in the near term.

To initiate discussions, WHO presented a proposal to focus the meeting on the following three metrics as the primary focus for the development of WHO technical guidance:

1. **Moderate-to-vigorous intensity physical activity:** MVPA is seen as a primary metric given it is the main aerobic physical activity recommendation for adults and older adults.
2. **Step count:** Step count is seen as “forward-thinking”, anticipating future WHO global guidelines in 2030 that may support step-count recommendations based on new epidemiological data.
3. **Sedentary behaviours:** Measuring and reporting a metric for time spent in sedentary behaviour(s) is desirable although it was noted that current WHO Guidelines do not provide a quantifiable metric, instead recommending that people “limit amount of time spent in sedentary behaviours”. Developing a method and protocols for wearable device-based measurement of sedentary behaviours is included as “forward thinking” in anticipation of the 2030 guidelines providing more specificity.

Measurement of other aspects of physical activity, while of interest, were also presented to be outside the scope of this meeting because for a range of reasons (e.g. no current WHO guideline, current insufficient evidence, no clear definition, or insufficiently developed wearable device technology to accurately measure the behaviour). These included:

- light-intensity physical activity;
- vigorous intermittent lifestyle physical activity (VILPA);
- multicomponent physical activity (as defined by WHO);
- muscle-strengthening activities;
- specific activity types (e.g. swimming, cycling);
- sleep behaviours.

³ Definition based on WHO glossary: multicomponent physical activity can improve physical function and decrease the risk of falls or injury from a fall; it typically combines different types of exercise (aerobic, muscle strengthening, and balance training; examples include walking backwards or sideways or standing on one foot while doing an upper body muscle-strengthening activity, such as bicep curls; dancing also combines aerobic and balance components.

Discussion highlights

The group agreed that MVPA, step count and sedentary behaviour were good candidates for device-based measurement in national population surveillance systems in the near term. Delegates noted that primary studies and recent systematic reviews provide evidence of strong associations between low step count and both mortality and morbidity, and point to consistent, beneficial step-count ranges to improve health.

WHO clarified that there is currently no global, WHO-endorsed recommendation or policy position on step counts. However, it recognized that the evidence-base is rapidly growing, making step count a potential useful metric for population health systems to use for tracking, as well as to use as a simple communications tool to promote physical activity among multiple audiences. It was noted by delegates that step counts may be considered exclusionary as the metric does not reflect the diversity of movement experiences across the population. More specifically, it does not acknowledge individuals who are non-ambulatory and may move through other means, such as pushing a wheelchair or other forms of wheeling.

In discussing the step-count metric it was noted that there will be a need to carefully review and revise the language and public facing “messaging” used to discuss step counts to ensure people who are non-ambulatory are not excluded from population-level physical activity and sedentary behaviour messaging (as they often are). This exclusion contributes to a lack of identification with public health campaigns and may further marginalize these groups. Research from SCI Action Canada Lab is currently exploring this issue, highlighting the urgent need for inclusive approaches in both language and measurement. Future work both in developing any global (or national) recommendations based on step counts might explore using more inclusive terminology such as “step/push counts” or “step or step-equivalent counts” to better represent all forms of mobility.

Participants noted the difficulty and lack of consensus on defining and measuring muscle-strengthening and balance-training activities. WHO stated this agenda must be put to one side to not delay making progress on the proposed core set. The group noted that analytic advances using raw data and pattern recognition may facilitate reaching consensus on the measurement of strengthening activities as well as sleep characteristics in the future.

WHO acknowledged the strong interest in the measurement and population health monitoring of sleep behaviours and the potential for a wearable device to be a viable measurement instrument. However, this was excluded from discussion at the meeting and from WHO’s work on technical guidance until WHO has developed a formal guideline on sleep.⁴

Delegates noted that some countries do not use the WHO STEPS survey, and that some countries have developed 24-hour movement guidelines that cover physical activity, sedentary behaviour and sleep. WHO acknowledged this diversity and suggested that the development of consensus on the protocols for use of wearable devices would have

.....
⁴ WHO noted that some countries request additional items in STEPS to assess sleep, and a variety of questions are currently being used as, and that to date, there is no WHO recommended standard sleep behaviour questionnaire.

relevance to all countries regardless of the specific details of their national guidelines, as the guidance on wearable devices may be adapted to country context as needed.

Delegates questioned WHO's interest in the use of mobile phones to monitor physical activity and whether a mobile phone was considered a wearable device. WHO stated that while there was some interest in the potential mechanisms to access and use mobile phone data, it presented many challenges, including data protection. At this stage WHO was focussed on actual wearable (i.e., worn) devices and future meetings would explore other digital technologies such as mobile phones.

It was noted that WHO STEPS includes adults aged up to 69 years and that it was important to also monitor physical activity in those aged over 70 years. WHO noted that some countries extend the age range.

Delegates questioned whether GPAQ would continue to be used or be replaced by a wearable device within the WHO STEPS, and if so, how would countries transition? WHO suggested that GPAQ may be replaced by a streamlined set of questions related to other aspects of physical activity and sedentary behaviours such as the contexts (or setting) related to the behaviours, thereby allowing the wearable device to become the primary instrument quantifying the behaviours. However, WHO also noted that for some countries, GPAQ would likely remain the primary instrument until wearable devices were an affordable and feasible alternative. WHO also noted that countries will require clear guidance and protocols on how to transition from current practice and that any replacement questions in STEPS would need to be developed and validated. WHO presented that the long-term vision was to get all countries using wearable devices and that countries are only at the start of this journey. Further discussion on how this process would unfold (and timelines) was deferred to a later session/Day 2.

Case study of use of wearable devices in Canada

Dr Colley provided an overview on the use of device-based measurement within the Canadian Health Measures Survey (CHMS). Dr Prince Ware outlined the use of those and other data by the Public Health Agency of Canada in its reporting and surveillance system, including the Physical Activity, Sedentary Behaviour and Sleep (PASS) Indicators.⁵

- CHMS includes 5000–6000 participants per two-year cycle, and produces nationally representative estimates of a wide range of reported and measured health determinants and outcomes. In addition to answering questions about physical activity, sleep and sedentary behaviour, CHMS participants wear an accelerometer around the waist.
- The CHMS commenced in 2007 and to date there have been seven cycles of data collection.
- For the first six cycles of the CHMS (2007 to 2019), the Actical (Philips Respironics, Oregon, United States of America) was worn on the waist and used to measure physical activity using a “waking-hours wear” protocol for seven days. Raw Actical data were

⁵ <https://health-infobase.canada.ca/pass/>

analysed using published SAS programmes (**Accel+ | Healthy Active Living and Obesity Research Group**).

- Starting in cycle 7 (2022–2024), the CHMS transitioned to a waist-worn Actigraph device to measure physical activity using a 24-hour wear protocol for seven days. Actigraph data will be processed and analysed using a publicly available R statistical package (to be released 17 October, 2025).
- To better understand trends over time, a cross-over study was conducted whereby 100 participants wore both the Actical and Actigraph on their waist for seven days. This work was undertaken to help assess the extent to which changes in physical activity levels at the population level are explained by the device and protocol change.
- Results from the 2018–2019 CHMS showed that approximately 50% of the population met recommended levels of physical activity based on waist-worn accelerometers.
- Prevalence for females was lower than males, especially in youth, and activity was lower with age for males and females. The prevalence data (CHMS 2007 to 2019 combined) showed a reasonable correspondence between prevalence of obtaining 7500 steps per day and prevalence of those meeting the criteria of 150 minutes MVPA per week (that is 53.4% of adults aged 18 to 79 years met the step target while 43.0% met the MVPA target. The data also showed that more than 50% of steps were accumulated during light-intensity physical activity.
- It was noted that the sample size for the CHMS survey design is such that device data can only produce national estimates, while other, larger questionnaire-based surveys (e.g. Canadian Community Health Survey) have the larger sample size required to produce subnational estimates.
- Another observation in Canada is the difference in results on sedentary behaviour from self-report and device-based datasets – this is because the sedentary behaviour recommendation (≤ 8 hours per day) adherence is defined differently based on the measure: for the device it is ≤ 9 h/d and for self-report questionnaire data it is ≤ 7 h/d (6). More details on the CHMS are available.⁶

Dr Colley shared details of current work at Statistics Canada looking into the willingness of individuals to upload personal data from their own (consumer) device. The “Share your Own Data” (SYOD) Study is currently consulting data users, clinicians, researchers, and the public. This would involve participants completing an electronic health questionnaire (already available) and uploading their own physical activity data files from their commercial wearable device. It was noted that there will be significant challenges in overcoming the obtaining, processing, and harmonizing of raw data from the vast number of different commercial wearables.

Related to this work, Dr Colley noted that Statistics Canada may consider adding a question to the health survey to ask respondents whether they have a personal wearable device,

.....
⁶ <https://www.statcan.gc.ca/en/survey/household/5071>; <https://www150.statcan.gc.ca/n1/en/pub/82-003-x/2010001/article/11066-eng.pdf?st=X-cUQGWL>; <https://www150.statcan.gc.ca/n1/en/pub/82-003-s/2007000/article/10361-eng.pdf?st=G56leV0a>; <https://www.statcan.gc.ca/en/survey/household/5071>

what brand it is, and whether they would be willing to share their own data (i.e., with the Canadian government). Data from these questions would help explore the potential biases associated with using voluntarily contributed data from consumer devices.

Discussion highlights

Discussion of the Canadian work included many questions on logistics related to sample selection, potential biases, and the use of sample weighting to address these issues. Key questions included the response rate of participants, which was reported to be ~40% for the first six cycles and lower for cycle seven. Reasons for the decline were discussed and included erosion in government trust and surveillance programme factors that make it more difficult to establish relationships with individuals in the way you would in clinical studies. Further details were requested on the potential bias in the sample and whether information on socio-economic status was known. Data loss due to required “wear days” and “wear hours” was discussed and Statistics Canada indicated this is a difficult balance between perfection and preserving as much data as possible to ensure adequate sample size.

WHO highlighted that Canada provided a strong example of a country adopting wearable devices in population health surveillance to measure physical activity and conducting a crossover study alongside a method change. WHO noted there would be interest from other countries to learn from the Canadian experience as well as from other countries where devices had been used in national surveys. WHO noted that Canada does not currently report on a step-based metric in their surveillance system and that both WHO and Canada were interested in better understanding the best metric to use for physical activity monitoring and how it might align with potential step count recommendations. WHO is also interested in exploring the potential of using data provided from consumer devices to compute national estimates and would like to follow up on this with Statistics Canada.

Session 3: What is this device going to do? What must it have, and what is nice to have? (Dr Fiona Bull, Dr Rick Troiano, Dr Matthew Ahmadi)

This session focused on Meeting Objective 2, namely specifying the desired characteristics of a wearable device for use in population surveillance. The session was supported by a background document (an updated version of which is contained in Annex 3) which presented the issues in four tables across four areas, namely:

1. Sample selection, device administration and compliance incentives (noting that these issues were not the primary focus of this third expert meeting)
2. Wear protocols
3. Device properties
4. Processing protocols

Before the discussion, WHO clarified the following points:

- The task was not about selecting a current brand or model of wearable for use, but rather to use the best available scientific evidence and practical field experiences to collectively define the characteristics, functions, and protocol, that any brand of device might need to have (or exceed) to be suitable for use in adult populations in national surveillance systems.
- WHO is brand-agnostic and recognizes that some available devices would already meet and exceed the likely characteristics this task will identify. However, available wearable devices (both “research grade” and consumer) are considered too expensive for adoption and use in large-scale population surveillance systems, especially in low- and middle-income countries.
- For any issue where the evidence is not available or a clear decision was not possible, WHO requested recommendations and proposals on what is needed to address the knowledge gaps.
- The device specifications must support defining a wearable instrument that would be practical, feasible, and affordable – and that the “good enough” principle should apply – i.e., perfection should not deter some compromises when needed because cost of the device was a limiting factor to large-scale use in monitoring systems.
- The possibility to identify “higher” or “desirable” specifications and/or functionality (i.e., deemed beneficial but not essential) would be considered for the WHO guidance and be discussed further in consultation steps.
- Specifications should allow the device to be applicable (useable) for all adult populations – including older adults and/or persons living with disabilities. Where this is not the case, it should be noted and alternatives offered for certain populations.
- In addition, issues related to data transfer and data security must be considered during the development of the device specifications and protocols for use

To initiate discussion, WHO outlined the tables provided in the background document (see Annex 3) and highlighted three issues on which WHO already has an initial position based on input obtained during prior expert meetings. These were the following:

1. **Wear location:** (Table A3.2), The proposed location is the non-dominant wrist as it is the most practical and acceptable location for use in a general population health-monitoring systems context – a position reached based on evidence from field experience and recommendations from experts from two previous meetings.
2. **Number of axes:** The device should include triaxial accelerometry (Table A3.3) – a position reached based on this now being low cost and standard in most devices;
3. **Access to raw signal data:** (Table A3.4) Seen as a non-negotiable core requirement based on the need to provide maximum access to data and control of data processing and analysis, this requirement provides the greatest flexibility and allows retrospective reprocessing of data should algorithms or metrics change or be added as knowledge and analytic techniques advance. This also would enable harmonization across device/settings if required.

Delegates discussed these three issues and started a review of all remaining issues presented in the tables.

Discussion highlights

During general discussion, the following points were made or reaffirmed:

- The group agreed with the WHO suggestions on a wrist worn device, triaxial accelerometry and access to raw data.
- It was noted and agreed that requiring raw data had implications for data storage and battery requirements of the device and that optimizing how the data were stored and transmitted to minimize cost and burden on data collection personnel in the 'field context' would be essential and required discussion.
- The group discussed the minimum number of wear days for data collection and the minimum number of days and hours per day required for inclusion. The debate was nuanced in terms of what the future guidance and protocols might state a country "should" collect versus what is deemed necessary ("acceptable") for use in data analysis given that balancing response and compliance rates can be challenging.
- As current WHO MVPA guidelines are stated using a "per week" timeframe, it was agreed that the protocol specifications should state that data be collected for a minimum of 7 days (see Table A3.2).
- Delegates also agreed that countries could be encouraged to collect data over additional days (i.e. more than 7 days) to enable a period of familiarization with the wearable device. Although this would add an additional burden, it was suggested that WHO guidance should indicate that countries could decide to collect data for more than 7 days within their country context.
- It was noted that for those countries that have national guidelines on physical activity that use a 24-hour time frame, 7 days data collection should still be recommended so that the population estimate can best reflect habitual patterns.
- However, the discussion did not reach agreement on the criteria for how many days of data collection from an individual should be stated as sufficient (i.e., a minimum) for inclusion in analysis, suggesting that it should be set to minimize random error that occurs due to day-day fluctuations.
- The group discussed the protocol for the number of wear hours per day with possibilities ranging from: 24 hours, or "at least 18 hours", or "at least 10 hours" per day, or a proportion of each hour of the day. As WHO is not providing guidance on the measurement of sleep at this time, it was not necessarily a prerequisite to state 24 hours per day wear time. This was however seen to be a practical approach, as 24 hours per day would mean the wrist-worn device would not need to be removed.
- Overall, the majority of the group accepted 24 hours/day but as the issue was not finalized. Table A3.1 records several options.

- It was noted that decisions on the majority of the device properties and specifications are highly interconnected, and many decisions affect sensitivity – and importantly – device cost.
- During the discussion of the remaining issues in Tables A3.2.4, delegates raised issues related to sample selection and device administration (these were originally contained in a table provided in advance of the meeting to delegates, but not discussed at the meeting as they were beyond its scope). This included questions on how devices would be delivered to participants (in-person or via post, as used in some countries), and how the device would be returned (in-person or via post or other pick-up points – as used in some countries). WHO stated that currently there is no fixed protocol, and that recommendations on these practical details needed to be developed. It was also stated that at this time, WHO considers that the device would most likely be provided in-person during the WHO STEPS (or equivalent) survey completion.
- WHO noted that more in-country studies were needed to identify and test practical and feasible ways to efficiently retrieve the devices within surveillance data collection systems such as WHO STEPS or equivalent national approaches. In the Canadian CHMS example, devices are provided in person and participants are provided with a prepaid envelope to return the device by post after 7 days.
- Delegates were reminded that WHO provides some technical support to countries to assist in the planning and implementing a WHO STEPS survey but that countries themselves undertake data collection and data analysis, and retain data ownership. Countries choose to share their data with WHO for the purposes of developing global estimates and for international comparisons. WHO does not own any country data and this will apply to all physical activity data collected with any wearable devices.
- In this session consensus was achieved on the following:
 - Wear location – non-dominant wrist.
 - Use of a triaxial accelerometer.
 - Collection of raw data.
 - Request 7 days' wear time, with options for countries to extend the number of days.
- The session did not complete a review of all remaining items in Tables A3.2.4 and concluded with agreement for a small working group to continue to review and annotate issues remaining across these tables, and to report back on Day 2. All delegates would have opportunity to review and comment, including on the applicability of suggestions for use in older adults.

Session 4: Validating, storing, analysing, and harmonizing data (Dr Aiden Doherty, Dr Manos Stamatakis, Dr Kerstin Bach, Dr Marco Giurgiu, Dr Sarah Keadle, Dr Charles Matthews)

This session focused on Meeting Objective 3 and was divided into two sections. Part 1 provided the opportunity to learn from experience using wearable devices with different protocols and analytic methods in large prospective cohort studies across multiple countries. Part 2 focused the discussion on the evidence on the validity of wearable devices, and on the limitations of current study designs and evidence base, and what future actions, including research, were needed.

Part 1: Dr Doherty, Dr Stamatakis, Dr Bach

Part 1 focused on sharing and comparing current open-source machine learning methods and cut-point methods; and learning more about work to develop and evaluate analytical code using self-supervised machine learning. The three speakers and their work on open-source methods were:

- Dr Doherty – OxWearables (from Oxford University) as used in the UK Biobank study (2013–2015), China Kadoorie Biobank (2020–2021), English Longitudinal Study of Aging (2021–2023) and the Chinese Health and Retirement Longitudinal Study (2020–2021).
- Dr Stamatakis – ProPASS methods developed for the Prospective Physical Activity Sitting and Sleep Consortium (ProPASS).
- Dr Bach – Self-supervised learning models as used in Trøndelag Health Study (HUNT4, 2017–2019).

Dr Doherty – Key points from presentation

- Imperfection should not be a barrier to making a transformative leap to use of wearables to measure physical activity and sedentary behaviours.
- Current available analytic methods are good enough for today.
- Prioritize raw data – data storage and processing of large amounts of raw data is not a problem given current computing capability.
- Open-source analysis tools are available and ready to use.
- Results from raw data analysis from wrist-worn Axivity (used in the United Kingdom of Great Britain and Northern Ireland) and Matrix 003 (used in China) provided comparable outcome data (data from forthcoming publications were shown) (7). Step-count results were found to be very similar, however the results for MVPA showed some discrepancy and raised some concerns which warranted further investigations.
- The OxWearables algorithm had better accuracy than the Verisense algorithm for producing step counts (12.5% vs 63.5% mean absolute percent error in free-living

environment, 16.6% vs 17.3% mean absolute percent error in laboratory environment). More details are available (8,9).

- Given MVPA is more challenging to estimate using machine learning methods, the Oxford analysis has reverted to using a cut-point approach to classify MVPA and light-intensity physical activity.
- More validation is desired, but to date, results from independent small samples with observed measure of steps are consistent in showing acceptability for the step-count algorithm.
- Globally, there is a need to reduce protocol heterogeneity and to improve the methods used in validation studies.
- More publicly accessible datasets are becoming available, and this could help improve validation methods.
- If different wrist-worn devices are used by countries, harmonizing data to produce step counts from different wrist-worn devices is unlikely to be a problem as data collected from different devices are highly similar (i.e., results from Axivity AX3 vs Matrix 003).
- Results from analysis of the UK Biobank study (10) have shown that 7 days' worth of data collection is sufficient and that more days are unnecessary.

Dr Doherty concluded by reminding the group that epidemiological studies with device-based data have shown stronger associations between physical activity and health outcomes than self-report measures, so it is time for this transformational leap in surveillance methods.

Discussion and Q&A highlights

- Although advances were acknowledged, delegates agreed that modelling MVPA remains difficult and that more (and better) validation with greater harmonization of methods is required.
- It was agreed that MVPA signal data patterns are diverse and that developing machine learning models for MVPA is therefore complicated, and that walking has less variability and can be classified more easily. It was strongly agreed that addressing this to make progress requires specifically more MVPA focus, development and evaluation.
- Discussion on the differences between dominant versus non-dominant wrist-wear location noted a slight advantage for non-dominant location; and this location was agreed as the preferred specification. As some individuals may decline to use this location, it was recommended that protocols require that the wearer's preferred location (i.e. which wrist) be recorded. Recording the cases where non-dominant wrist was used could help inform the analytical model and/or help explain outliers.
- Discussions questioned what evidence supported the use of the OxWearables algorithm with older adults and it was generally agreed that more evidence was needed on the use of wearables in older adult populations and studies with larger sample size.

- However, additional evidence was shared from the Mobilize-D Study which recruited 2400 older adult participants across four main clinical cohorts: Parkinson's disease; multiple sclerosis; chronic obstructive pulmonary disease; and proximal femoral fracture. In brief, the Mobilize-D study developed and validated the MobiGap pipeline (algorithm) focusing on detecting gait and characterizing daily living gait in people with neurological disease (Parkinson's disease, multiple sclerosis) or altered gait patterns for some other medical reason. However, the study has not included an assessment and metrics on outcomes of physical activity in their analysis pipeline. Further details on the validation studies conducted in the project are available (11).
- Delegates agreed that there was a need for more evidence on, and improved algorithms for, application with older adults and those with mobility impairments.
- Delegates generally agreed on the need for a mechanism(s) to support more researchers to share data to encourage and facilitate more cross-validation of algorithm work.

Dr Stamatakis – Key points from presentation

- ProPASS is a consortium of prospective cohort studies using wearable devices with a focus on standardizing and harmonizing measurement and validation, developing data processing tools, and building capacity globally in epidemiological research on physical activity, sedentary, sleep and health.
- The consortium aims to increase consistency of methods and diversity of populations represented in epidemiological evidence with a primary focus on building and supporting epidemiological evidence from low- and middle-income countries.
- ProPASS was developed to address the lack of standardized methods for fieldwork/ measurement/data processing; lack of data-sharing mechanisms; address the design of devices with updated technology at an affordable price; and to support the transition of global and national physical activity guidelines from being self-report based to wearable-device based.
- ProPASS has developed the DataSHIELD platform – a novel data-sharing mechanism that allows data to remain on study servers in each country, and yet enables data from multiple cohorts to be securely accessed and pooled, and analysis to be performed remotely. This addresses data security and protection requirements and avoids data transfer across borders.
- ProPASS has developed a standardized data processing tool, ActiPASS (University of Sydney), which is brand-agnostic, and will soon be dual-placement compatible (wrist or thigh).
- ActiPASS can measure many aspects of physical activity, sedentary and sleep, and most components are either validated, or in the process of being validated.

- ActiPASS Pro is also open source and the evolution of ActiPASS is such that it would allow incorporation of any algorithm; it therefore could be considered as a way to support WHO's work.
- The Sydney team has also developed a two-stage wrist accelerometry classifier that classifies broad activity type and then applies universal cut-points for walking-based activities. This classifier has been applied to data from the UK Biobank study, as well as the United States of America's National Health and Nutrition Examination Survey (NHANES) and has been adopted by other research groups.
- Evaluation results to date show better accuracy estimates for moderate and vigorous intensity physical activity than for light physical activity, and for sedentary behaviour, but all activity classes perform considerably better than any questionnaires.
- Classification of activity type and posture performs in a very comparable manner whether the device was worn on the wrist or thigh, based on ProPASS comparison studies.
- Dr Stamatakis noted that WHO will potentially need support from a central hub to assist with early stages of country use of wearables and support for data analysis – at the very least at the beginning to help front-runner countries. He noted that ProPASS may be able to offer support and also to share protocols and potential devices to support studies in low- and middle-income countries.

Discussion and Q&A highlights

- The need for independent testing and evaluation of algorithms was discussed and strongly supported, leading to numerous suggestions and proposals for the type of future research needed to advance this field.
- It was agreed that the available evidence (on wearables) represents a substantial improvement in the accuracy for core metrics of physical activity well beyond what is available from current survey methods. Many (most) in the group considered that currently available evidence on validity was “good enough” for the global agenda of WHO to proceed to promote the use of wearables in population surveillance systems.
- However, it was also agreed that stronger evidence was needed to demonstrate the accuracy and precision of algorithms in more diverse contexts to show that “wearables can work” and to build trust and interest in using wearables in more countries, particularly in low- and middle-income countries.
- Delegates discussed how to decide when to change or update the analysis models (both the prediction and the classification algorithms). The main issue was: how much improvement in a model is needed before a change is warranted? Suggestions for candidate criteria included:
 - F1 (for MVPA and sedentary time);⁷
 - mean absolute percent error (for step counts metrics);
 - considering Root Mean Square Error (RMSE).

⁷ The F1 score is a machine learning metric that provides a balanced measure of a model's precision and recall.

- All agreed that making changes or updates to analytical methods (e.g. machine learning algorithms) should be carefully considered, and short-term updates should be resisted. The discussion on validation study methods highlighted agreement in the need for more sharing of protocols on how to collect data and annotate the data; this prompted detailed discussion that was postponed for later/Day 2.

Dr Bach – Key points from presentation

Dr Bach introduced the self-supervised learning models being developed and applied in the Trøndelag Health Study (HUNT).

- HUNT4 data is based on 35 000 participants providing 6 full days of accelerometer recording using devices worn on the thigh or lower-back – this selection was based on earlier decisions and study design.
- The self-supervised learning approach to developing a model uses a large volume of unlabelled data as well as a smaller amount of labelled data – the process is analogous to large-language models such as ChatGPT.
- The self-supervised learning approach does not require the implicit learning of features. During training, the model develops an internal representation of the data, which is then used to classify activities. Due to the fact that unlabelled data can be used, larger amount of data is available which improves model performance.

It comprises two core steps:

- **Upstream tasks** include data collection, sensor calibration, pre-processing, and feature extraction. Raw accelerometer signals are cleaned (e.g., noise reduction, signal filtering), segmented into meaningful time windows, and transformed into features such as step counts, activity intensity, or frequency patterns. These steps ensure the data is standardized, labelled where possible, and suitable for model training.
- **Downstream tasks** focus on applying and refining the model to achieve real-world objectives. Once trained, the model classifies activities (e.g., walking, sedentary time), estimates key metrics (e.g., time in moderate-to-vigorous physical activity), and adapts to new data through self-learning methods such as *mask-and-swap* – a technique where parts of the input are hidden or replaced to test robustness and improve generalization.
- The masking of spectrograms turned out to be the best approach to teach the model data characteristics and classification of human activity. We also experimented with single and cross-sensor data (back and thigh data) which looked promising (12).
- All data are available on GitHub, and more information is available.^{8,9}
- Having access to more raw data will allow the learning of more comprehensive features.
- Existing data can be used to build models.

⁸ Open code and dataset for supervised learning: <https://github.com/ntnu-ai-lab/harth-ml-experiments>

⁹ Open code and trained SSL models: <https://github.com/ntnu-ai-lab/SelfPAB>

Discussion and Q&A highlights

- It was noted that self-supervised learning was also used in developing the OxWearable algorithm but using different data and a different approach (data augmentation versus data masking); the ActiPASS approach was noted as not using self-supervised learning.
- Self-supervised learning is a way to improve algorithms that may lead to significant improvements, especially for classification of MVPA.
- Broader implementation of self-supervised learning will require the collection of more data to train the model.
- In addition, better labelled data are needed, particularly for the downstream step and to help improve classification and analysis of MVPA.
- The HUNT4 method is not currently ready for use with wrist data as it is trained on back- and thigh-worn devices; in contrast, OxWearables and ActiPASS are both ready to use if deemed acceptable.

Further discussion on the acceptability criteria and adequacy of the OxWearables and ActiPASS model validation was earmarked for a later session, including more discussion on their readiness and generalizability for use in other countries, especially low- and middle-income countries, and in older adults.

Part 2: Dr Giurgiu, Dr Keadle, Dr Matthews

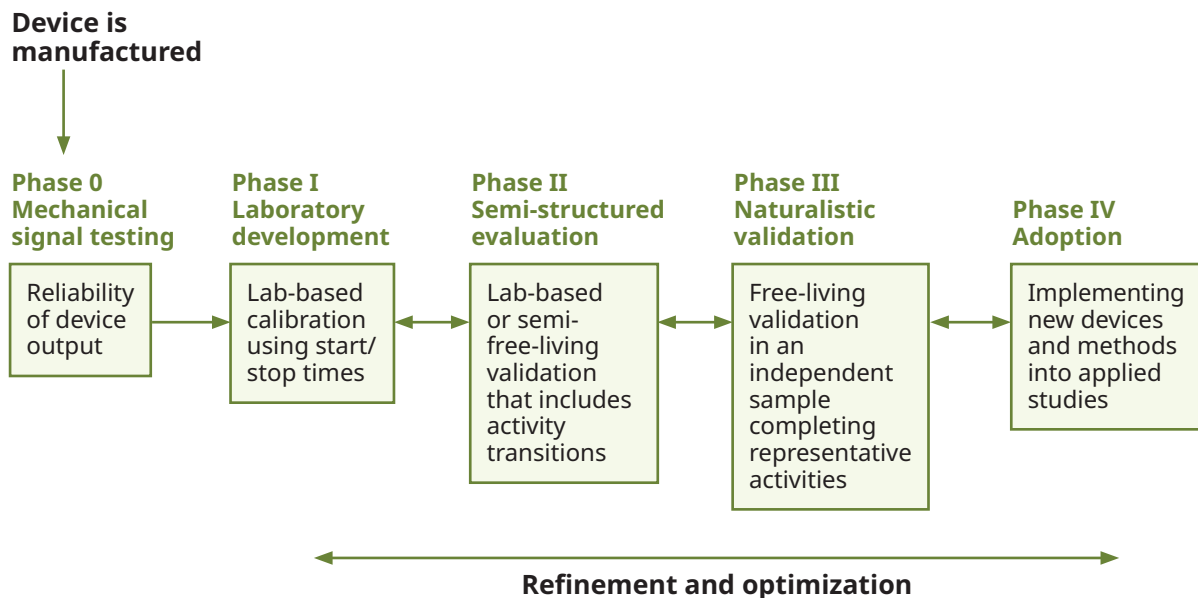
Drs Giurgiu, Keadle and Matthews led a session discussing current methods of validation, status of evidence on the validity of wearable devices, and limitations with study designs and gaps in knowledge. Currently, the evaluation of wearable devices in the physical activity field faces a major challenge due to non-comparability, as methods and procedures to develop and calibrate algorithms vary widely across studies, and rigorous evaluation of the algorithms in real-world settings is limited. The primary aim of the session was to identify what work is needed to strengthen evidence on validity of wearable devices, and how to improve study methods to build better evidence with a specific focus on step counts, MVPA, and sedentary behaviour.

Dr Keadle opened by providing an overview of the “Framework to evaluate wearable devices”(13) – an existing validation framework for wearable device-based measurement of physical activity. A summary of the recommended steps for algorithm development and verification was outlined (see Fig. 1). Additional key points were:

- Establish reference measures and procedures, including:
 - Clearly define behavioural metrics and corresponding primary reference measures (e.g. direct observation for sedentary behaviour and step counting).
 - Demonstrate that the selected secondary references (e.g. thigh-worn accelerometers for sedentary behaviour; or StepWatch, an ankle worn device for step counting) are proxies (traceable) for primary reference measures that are better suited for data collection in free-living environments (also referred to as field-based reference measures).

- Develop and calibrate the algorithm in laboratory and/or natural protocols.
- Evaluate the performance of algorithms in real-world settings in independent samples with representative activities using secondary reference measures (i.e., Phase III studies).

Fig. 1. Overview of a phase-based framework to evaluate body-worn devices to assess physical behaviour (13)



Dr Matthews reflected that to date, the physical activity measurement field has mostly relied on showing associations with health outcomes to justify confidence in different measures, but today accuracy of the values obtained from algorithms (e.g. minutes of MVPA) is more important than previously thought, particularly when the focus is on the development of physical activity guidelines; and when monitoring compliance with the guidelines relies on specific thresholds for a given metric.

Following a useful discussion on semantics, the group agreed that the term “evaluation” would be a much better to use than “validation”, as evaluation provides positive or negative results based on clear criteria while validation implies an affirmative result. However, introducing this shift in terminology will take time and it was noted that “validation” is established terminology in data and computer science and more commonly used.

Several key concepts were introduced that could help advance the science by increasing transparency and reducing heterogeneity:

- Metric definition – clear operational definitions that can be applied across studies (e.g. sedentary behaviour, steps, MVPA).
- Unit calibration – the process of testing the reliability and validity of raw-data signal generation in electronic/mechanical settings – this should be conducted in controlled laboratory settings (e.g. a shaker table).

- Value calibration – the process of developing an algorithm to estimate activity energy cost or type (conducted in controlled laboratory settings with prescribed activities).
- Criterion/primary reference measures – an accepted gold-standard method for measuring a particular metric of interest (e.g. video-recorded direct observation for step counting or sedentary behaviour).
- Field-based criterion measures (e.g. ankle-worn stepwatch or a thigh-worn device for step count and sedentary behaviours compared with direct observation).
- Traceability – the established relationship of accuracy and precision between a primary (criterion) method and a secondary method – establishing and communicating traceability ensures that algorithms are anchored to and tested against a common (clearly stated) standard.
- Commutability – refers to how well the calibration procedures mimic behaviour in real-world settings for training algorithms to generate accurate results in verification testing.

Dr Matthews summarized that to improve performance, evaluation studies must:

- establish clear operational definitions for behavioural metrics of interest;
- develop and adopt traceable reference measurement methods for real-world use;
- seek to identify algorithm development procedures that “work” when the algorithm is applied in real-life settings (i.e., commutable procedures).

Dr Giurgiu presented the key findings from the systematic review of validation studies for device-based measures published 1970–2021(14):

- Out of 967 studies, nearly 80% had high risk of bias.
- Of note, only 11% of the validation studies were conducted among adolescents.
- The most commonly validated devices (i.e., the top 50) evaluated intensity-related metrics such as steps or energy expenditure, while very few devices evaluated metrics from different domains (e.g. intensity, sleep, and posture).
- Overall, the variability between validation methods used in studies makes it nearly impossible to evaluate differences among devices and led to development of the Wearable Landscape Network (<https://wearable-landscape.info/vision>).
- The network was created to bring together researchers and industry partners (currently 153 researchers and industry partners from across 36 countries) to conduct collaborative activities related to wearables. The website is intended to serve as a central point of contact for overviews and exchanges of tools, projects, and topics in the field of wearables.
- Current work includes conducting a “living review” to continually update the systematic literature review (albeit currently operating with no external funding); and Delphi studies to work on a checklist for reporting relevant information, as well as developing a quality-assessment tool to judge the risk of bias of evaluation studies.

Discussion and Q&A highlights

- Delegates and WHO questioned how well the two open-source machine learning models presented will generalize to population groups outside those represented in training (calibration) and validation studies.
- Overall, it was agreed that generalization (of algorithms and use across populations) may be more robust and acceptable for step-count metrics and possibly also for sedentary behaviour, but the classification of MVPA was more uncertain. This was considered to be because the most prevalent MVPA activities (e.g. occupational versus housework versus leisure) are more variable in type and movement pattern, therefore additional research is needed, particularly in low- and middle-income countries.
- More evaluation evidence is needed but a key barrier to progress is obtaining funding for performance-evaluation studies.
- Collaborative projects and sharing of already collected data will facilitate some of the evaluation work identified as needed and should be explored and supported. Therefore, there is great value in identifying what data are currently available and discussing how to collate and pool data to help answer some of the emerging questions and evidence gaps (further in-depth discussion deferred for Day 2).
- Discussion explored the possibilities of collating and pooling data to develop the MVPA analysis (further in-depth discussion postponed for Day 2).
- Discussion raised the need for more data from low- and middle-income countries and the need for a multi-country studies with wearable devices in low- and middle-income countries (further in-depth discussion deferred for Day 2).
- A delegate suggested it might be beneficial to establish a pre-registration system for evaluation studies (including protocol and analytical approach) to reduce reporting bias in publications and contribute to advancing scientific methods and knowledge generation.
- It was noted that measures of sedentary behaviour and step count assessed by a wrist-worn device could be evaluated against accepted field-based criterion measures (i.e. thigh-worn and ankle-worn devices, respectively) in new studies or using existing datasets.
- WHO was asked if it wanted to develop a new wearable device, and if so, would it need to be validated or, if the specific components of the device were well known (and tested in other devices), would the manufacturers' report on validation/performance be accepted? There was some debate about this, but it was considered that broadly for the component parts inside a device, the manufacturer's product specifications would be acceptable. However, if a new device was developed, the overall performance would require evaluation to be deemed acceptable.
- The group was reminded that population-based surveillance metrics are more robust in relation to individual measurement because the goal is to estimate population means, which are less variable than assessing behaviour at individual level.

The session finished with a rapid “white board” discussion to collect suggestions on what are the next steps to improve evaluation studies. The points raised included the following:

- Study setting
 - Start with higher quality, laboratory-based protocols and include free-living protocols with primary or traceable secondary reference measures.
- Sample diversity
 - Conduct studies with greater diversity of sample populations and diversity in gait characteristics including children, adolescents, older adults, and populations from low- and middle-income countries.
 - Collate and pool existing data and develop a repository for validation data.
- Reference measures
 - Identify, agree and use traceable wearable data as a secondary measure to direct observation (e.g. thigh-worn devices for posture).
- Analytical approach
 - To reduce the variation in methods, there should be recommendations on analytic approaches, e.g. MAPE, Bland-Altman, equivalence testing.
- Independence
 - To develop trust, conduct more independent analysis of models (i.e. , in a dataset and team that differ from those where the algorithm was originally developed).
 - Develop an independent training dataset (for use to test and “benchmark”).

Day 2

The Day 2 agenda was revised according to progress made on Day 1. It was agreed to repurpose Session 5 and 6 to continue discussion on the specifications of device attributes, protocols and evaluation evidence. As such, the session on “other data needs” was deprioritized and moved to later in the day, if time allowed. In addition, it was agreed to allocate time during Day 2 for WHO to outline: (1) how it would start pilot-testing (demonstrating) the use of wearable devices in a country national surveillance system; and (2) what consultation should be undertaken as part of the development of the planned WHO technical guidance on wearable devices.

Session 5: Follow-on work on specifications of device attributes, protocols and data analysis – continued with report-back from sub-group (Dr Aiden Doherty)

The subgroup provided an overview of their discussions and presented an updated version of Tables A3.2.4. They outlined their initial recommendation on each issue, which supported a group discussion that proceeded to systematically review these issues, namely: wear protocols (Table A3.2–4); device properties (Table A3.3); and processing and analysis protocols (Table A3.4). No discussion was held on sample selection and device administration (Table A3.1).

For ease, Table 1 below present a summary of the decisions made on each issue. For full details of the decision (or no decision), brief rationale and any additional notes of the discussion made during the meeting, please refer to full versions of the tables in Annex 3. Please note the summary below and Annex 3 reflect the position reached at the completion of the post-meeting electronic review of the meeting report.

Table 1. Summary of decisions on device specifications and protocols

Sample selection and device administration summary

1.1 Sampling (and subsampling) method	These issues were not the focus of this meeting and not discussed in any detail and therefore no recommendations.
1.2 Recruitment method	
1.3 Adjustment for sampling bias	
1.4 Device administration	
1.5 Compliance strategies	
1.6 Collection of data related to physical activity	
1.7 Collection of socio-demographic data	
1.8 Collection of any environmental/ geographic data	
1.9 Training for data collectors	

Wear protocols summary

2.1 Wear location	Wrist worn – non-dominant AND Recommended to record the wear location for each responder as some individuals may decline to wear on non-dominant wrist
2.2 Wear time	Wear days: Minimum of 7 days / Optional: collect >7 days Wear time: No decision
2.3 Device orientation	Specification of non-dominant wrist

Device properties including sensor specifications summary

3.1 Number of axes	Minimum: Triaxial
3.2 Dynamic range	Minimum: +/- 6g
3.3 Resolution	Minimum: 12 bits – 3 milli-g (milli-gravity units)
3.4. Sensitivity	A Least Significant Bit (LSB) of ~0.003
3.5 Noise Floor	Minimum: 1–5 milli-g
3.6 Temperature sensitivity	Minimum: operational range should be minus 20°C to plus 50°C
3.7 Sampling frequency	Minimum: 30Hz Optional – 40Hz or higher
3.8 Memory capacity	Minimum: Memory capacity must be sufficient to support data collection at the specified number of days, hours per day and sample frequency AND Raw accelerometer data must be stored in a non-volatile was (to avoid data loss should battery life end)
3.9 Battery life	Minimum: Battery needs to be sufficient to support the required data collection days and hours per day at specified sample frequency AND Preferably the device should not require the user to re-charge during the data-collection period AND If the device does require re-charging there must be mechanism(s) to protect the data and avoid user access or user download

3.10 Battery recharging and replacement	Minimum: Device must be re-usable and re-chargeable by central coordinator to enable re-use within the surveillance programme AND Devices must be re-chargeable at scale (i.e., with use of a docking station for large batch recharging)
3.11 Feedback to user	<u>Provision of feedback on physical activity/sedentary:</u> No feedback should be provided to user on physical activity (steps, minutes or other metrics) or on sedentary behaviours/time <u>Provision of other feedback or device uses:</u> No decision
3.12 Use of multi-sensor approach	Minimum: (i) Battery health and (ii) Temperature Optional – Temperature and battery health sensors can be switched on/off depending on country context. <u>OTHER SENSORS [to assist analysis/ other purposes/research]</u> (i) Heart rate (via PPG) – no decision (ii) Global positioning systems (GPS) – exclude (not required) (ii) Gyroscope – optional – as this has significant impact on battery requirements
3.12 Waterproof and dustproof and intended for prolonged skin contact	Minimum: (i) Water resistant (IPx8) (ii) Dust tight (IP6x) (iii) materials certified as hypoallergenic and biocompatible and intended for prolonged skin contact
3.13 Band fitting	Minimum: Wrist band must be adjustable, removable and cleanable or disposable (and replacement units available)
3.14 Data handling and transfer	No decision

Processing protocols	
4.1 Access to raw signal	Minimum: Raw accelerometer data must be accessible
4.2 Frequency filtering	Minimum: 20 Hz Butterworth filter
4.3 Epoch length	Maximum: 60 seconds Optional 10 sec or 30 seconds – not decided
4.4 Minimum bout length and interruption tolerance	<u>Use of minimum length of time (i.e. defining a “bout”):</u> Not required/recommended to use in data analysis. <u>Tolerance for interruptions:</u> No decisions
4.5 Identifying non-wear	No decisions
4.6 Minimum wear-time inclusion criteria	No decisions

Session 6: Advancing evaluation evidence (Dr Sarah Keadle and Dr Fiona Bull)

This session was repurposed to allow more discussion on issues arising from sessions 4 and 5.

Issue 1. Concerns on the evidence gaps in, and independence of, evaluation evidence of the currently available open-source algorithms to use for step counts, sedentary, and MVPA

Participants agreed that to improve the scientific rigour and to build trust in using wearable devices to measure physical activity and sedentary behaviours, more evidence was needed from testing algorithms on independent datasets (i.e. datasets not used to develop the algorithm). Secondly, it was agreed there was a need for more evidence from evaluation studies with statistical analysis conducted independently of the developers of the algorithm.

Dr Keadle shared initial progress made across Day 1 and 2 in collecting a preliminary list of known studies and researchers who were custodians of potentially relevant datasets for the above purposes. Delegates discussed and added to this draft list.

Delegates raised the issue of comparability of methods used for labelling and tagging of different physical activities. It was agreed that MVPA labelling should be improved and there is a need for consensus on how to label/define criterion data – particularly for MVPA, less so for sedentary and step counts as this was deemed more consistent across studies with less variability in behaviour. Despite the lack of consensus on operational definitions, some data supports that core metrics are labelled similarly in direct observation protocols

across independent research groups (15). Although it was agreed that existing datasets may have used very different labelling methods and therefore may be “messy”, they would still be useable for studies testing algorithms since the data (with its labelling differences/errors) would be consistent across the testing of different algorithms.

The group discussed optimal study design and methods to conduct independent, performance-evaluation studies for each of the core metrics of primary interest to WHO. It was agreed that a clear process was needed to oversee and conduct an independent evaluation. The project would require establishing a dataset, and analysis would be conducted independently from the algorithm developers. It was noted that the independent evaluator team should have sufficient skills to be able to handle the coding needs of running various algorithms on the contributed, possibly pooled data. The role of WHO in such a process was raised and seen to be important to provide support and justification.

The discussion concluded with the recommendation that WHO consider establishing an expert working group to undertake more detailed planning of the required research, including developing the proposed methods and identifying what resources and collaborators would be required. This proposal received positive interest from delegates and was seen as a something that could be commenced and completed in a relatively short period of time. Further discussion on establishing a global research consortium / working group was earmarked for Day 3.

In addition, the group discussed and proposed that a global collaborative effort was needed to identify, collate, and maximize use of all available “training datasets” – including, however, only those with acceptable methods and labelling of physical activity. Developing such a “pooled training dataset” could provide a global (central) resource and be available to research teams and developers to support testing and development of new and updated versions of algorithms.

Further, as part of future work, a “benchmarking dataset” could be created by using some of the pooled data and reserving it for use to test and “benchmark” algorithms. It was proposed that WHO could then “recommend” the use of any algorithms that had been independently tested and shown to meet set criteria for performance on the benchmarking dataset. This was seen as a longer term, but desirable, agenda.

Issue 2. Gaps in the evaluation evidence of currently available open-source algorithms in diverse populations, including low- and middle-income countries

In this discussion WHO returned to the question of whether the two currently available open-source algorithms were viewed as acceptable (and as such “ready to go”) for use in government health surveillance systems in countries other than those where they were developed. This issue was of particular concern to WHO because the integration of a wearable device into population surveillance systems would require countries to change from a known approach (i.e. a self-report survey) to a new and potentially more logistically challenging method using a wearable device. Therefore, WHO expected that

Ministries of Health (and their statistical agencies' data and relevant scientific experts) would want good evidence on the rigour and applicability of the proposed instrument, methods and data analyses; as well as potentially requesting evidence or use and results from a demonstration in countries similar to their own.

To explore the importance of this issue, details were requested about the testing populations used for the currently available algorithms. The details shared revealed that all were developed and tested in populations from high-income countries with limited studies in older people:

- OxWearables – developed and tested in populations from the United Kingdom of Great Britain and Northern Ireland, the United States of America, Canada, China, and underway in South Africa.
- Sydney 2-stage wrist classifier – developed and tested in population from Australia, the United Kingdom of Great Britain and Northern Ireland and the United States of America.
- ActiPASS Pro (wrist and thigh) – developed and tested in populations from Denmark, Sweden and Australia.
- Dr Bach's machine learning approach is being developed with data from population in Norway.

With this information, delegates confirmed there was a need to conduct more research with wearable devices in more diverse populations, and especially in low- and middle-income countries. Such work would help to evaluate applicability to other populations and contexts and establish more globally generalizable evidence.

For comparison, WHO noted that a similar situation occurred during the development and validation of the GPAQ instrument (and the International physical activity questionnaire – IPAQ). Both instruments were developed through global collaborations. For GPAQ, WHO established a GPAQ reliability and validity (evaluation) study involving nine low- and middle-income countries over approximately two years. The GPAQ consortia involved government and academic partners and Dr Bull was the chief investigator. More details are available (16).

The group discussed how to expedite a similar research programme with wearable devices, and the following suggestions were made:

- Explore potential opportunities within the ProPASS consortium and/or with countries that have an existing relationship with ProPASS (this included Tanzania).
- Explore opportunities to do evaluation studies using wearables in countries that are in discussion with WHO about conducting WHO STEPS in 2026–2028 and how conducting this research could be undertaken alongside – but not necessarily within – WHO STEPS data collection.
- Plan and conduct a few new “small studies” in interested countries using consistent – ideally standardized – protocols.

In terms of the some of the specifics; delegates suggested:

- Use thigh and ankle devices as practical reference measures for evaluating sedentary behaviours and step counts; and waist worn device for MVPA.
- Consider using photo/video criterion measures or a combined movement sensor and heart rate for MVPA, but it was noted that this method is a more complicated and labour-intensive protocol.
- Consider sensor-triggered (based on intensity and heart rate) ecological momentary assessments as a potential secondary reference measure.

Although it was agreed that the above research was needed, the session closed with a return to the main question of whether the current algorithms could be considered as “ready to go” for use in adult populations by countries. The majority of the delegates agreed they were ready for use in adults but not for older adults, and that they were more robust for step counts and sedentary behaviour than for MVPA. For the first two metrics, it was agreed that there was no reason to expect much variation in the performance of the algorithms across adult populations from different countries because the behaviours (and more importantly the signal) would be quite similar.

Delegates agreed that the current evidence on MVPA was not as strong. However some delegates put forward that the methods for MVPA should be considered “good enough” to use now as the device-based measure is still much better than the current self-report survey method, even if it is not perfect and there are evidence gaps.

After a lively discussion of the evidence on MVPA and need for more research (with both improved study protocols and more diverse populations and training datasets), the group proposed that a working group should be established to design the protocols and study design to conduct more testing of the existing algorithms; and that mechanisms to conduct a wider research programme involving more diverse populations, older adults and low- and middle-income countries should be sought to advance this agenda.

Session 7: Age-related considerations and sedentary behaviours (Dr Marla Beauchamp)

This session addressed Objective 6. Dr Beauchamp shared findings from an international consensus meeting on wearable technology for older adults held in Canada in November 2024 with 24 international experts (17). This meeting focused on older adults as a growing proportion of the population and discussed findings from the McMaster Monitoring My Mobility (MacM3) and Canadian Longitudinal Study of Aging (CLSA). Both these studies of older adults used a combination of wrist- and thigh-worn devices, with GPS to explore spatial aspects of physical activity and mobility, including active and passive transport and trips using contextual data from GPS. The meeting was held to address key issues related to using wearable devices in older adults, namely:

- slower walking speed;
- altered gait and use of mobility aids such as canes and rollators;
- applicability of current published validation studies and algorithms that have been predominantly conducted among younger healthier populations.

The consensus meeting aimed to develop recommendations on use of wearables in older adults to measure mobility across six domains, in line with the WHO's *World Report on Ageing (18)*, namely body posture, walking, physical activity, transport, life space, and transitions (e.g. from chair or bed). Within each mobility domain, meeting participants achieved consensus (>80%) on critical measures (out of 15–30 considered) for that domain resulting in a total of 20 core device-derived measures:

- **Physical activity (five core measures)** – Volume of movement (total activity/day or week); distribution and time patterns (bouts); type/mode of exercise (e.g. aerobic, strength, balance); activity intensity; step count.
- **Walking (four core measures)** – Step count; cadence; frequency and duration of bouts; walking/gait speed (max, mean).
- **Transitions (three core measures)** – Changing body postures (sit-stand-walk-lie); number of transitions (per day or hour); duration/length of time to transition.
- **Body posture (two core measures)** – Time spent in different postures; frequency, duration, distribution pattern of different posture bouts.
- **Transport (four core measures)** – Transport mode (e.g. walking, cycling, vehicle); time in active travel; distance travelled; trip frequency.
- **Life space (three core measures)** – Number of locations or trips; spatial areas and variation; time at home versus time out of home/outdoors.

The Canadian meeting also considered guiding principles for device-based measurement of mobility in older adults and reached consensus on many issues that overlapped with the agenda of this meeting, including:

- Devices should provide access to raw data and have sufficient battery for continuous monitoring.
- Data collection for 7 days is preferred for longitudinal studies with ageing populations.
- Data resolution should be as high as possible within constraints of battery and memory.
- Analysis should be with open-source software and using a population-appropriate validated algorithm.
- Derived measures should be validated or provide access to data and processing code to support replication/validity.
- Standardize reporting of methodology, using an appropriate checklist/guideline.

Dr Beauchamp also shared results from an evaluation study involving 746 older adults (mean age 73 years) in the MacM3 cohort, which used both thigh- and wrist-worn device data (collecting data for a minimum 10 hours over 7+ days). This showed better agreement between step count from the two wear locations using validated open-source algorithms (OxWearables for wrist and NiMBalWear for thigh) than with proprietary ones (TicWatch for wrist and ActiLife for thigh) (ICC 0.88 vs ICC 0.53).

Discussion and Q&A highlights

Several issues raised related to measuring step counts in older adults, including:

- Choice of device and wear location depends on which measurement(s) are of priority interest and the specific population of older adults.
- In older adults with neurodegenerative conditions that affect walking speed or arm movement, a wrist-worn device may misclassify walking as sedentary behaviour.
- There is a need for a clear definition of a “step” (e.g. “step” versus “stride”).
- There was concern about applicability and accuracy of step-classification algorithms that were developed among younger age groups being used for older adults.
- Delegates discussed that wrist-device measures can perform particularly poorly in unhealthy older adults and yet perform satisfactorily for healthier older adults. It was also noted that wearable data from (for example) younger adults with osteoarthritis and using a walker may also be poorly analysed despite their youth. In other words, how well the algorithm performs is not just related to age but also function and gait.
- Delegates discussed whether there should be more focus on disabilities and health conditions rather than simply on chronological age, as WHO STEPS generally does not survey adults over the age of 69 years. It was noted that 20% of the population have a disability and it is estimated that approximately half of those have mobility impairments; this population is more inactive and has a higher risk of disease.
- There was considerable discussion around the importance of algorithms being specifically tuned to those with mobility impairments, and the importance of including such individuals in training data sets to improve estimates for this population.

Session 8: What are the likely next steps following this meeting? (Dr Fiona Bull)

WHO outlined the potential next steps following this expert meeting. In particular, WHO was invited share its thoughts on how a research consortium might be established to help address some of the identified evidence gaps and to explain how a multi-country (including low- and middle-income country) research collaboration could be undertaken to help strengthen the evidence of feasibility of using wearable devices in diverse contexts.

To provide direction and identify key follow-up actions, Dr Bull outlined the following potential next steps:

- **Complete and publish the Montreal meeting report** – this would involve drafting and circulating the report to attending delegates to review and revise; finalize the report, complete WHO clearance, and publish on the WHO website; and promoting and sharing the report through a planned global (open/free access) webinar scheduled by the International Society for the Measurement of Physical Behaviour scheduled for 4–5 Nov 2025.
- **Draft WHO technical guidance document and share for consultation** – this would involve producing a first draft of the device specifications and protocols with an initial focus on adult populations. This draft would potentially be updated and expanded to include adolescent populations after completion of a similar expert meeting to address device specification for this age group in 2026.

An online consultation phase would be undertaken (potentially through webinars) between October and December 2025 (although this was noted to be ambitious and may extend into 2026). This would aim to include participants of this meeting and prior meetings; WHO technical units – including surveillance teams at WHO HQ and regional offices; the wider scientific community; Member States; and relevant industry stakeholders.

WHO suggested that the draft technical guidance document would be maintained as a “working document” while parallel research work was undertaken (see below).

- **Addressing evidence gaps through collaborative research** – WHO summarized emerging evidence gaps and the suggested lines for future research identified by delegates throughout the meeting. This included the need for more evidence generated through:
 - evaluation of wearable devices in studies involving low- and middle-income countries’ populations (referred to as a “nine-country study” in the meeting, while acknowledging that this figure was arbitrary but mirrored the development work undertaken for the GPAQ instrument discussed in Session 6);
 - evaluation of wearable devices with older adult populations and in populations with more diversity of functional ability and movement patterns;
 - further work to enhance the performance of algorithm classifications of MVPA, thereby strengthening evidence on the validity of wrist-worn wearable devices to provide a population-based estimate on meeting the WHO MVPA guideline;
 - independent evaluation of currently available analyses methods, including in particular the two open-source machine learning algorithms;
 - conducting national demonstration studies to test the feasibility of using wearable devices in population health monitoring systems in low- and middle-income countries (which could be done in countries interested in carrying out WHO STEPS);
 - the experiences and protocols on the use of wearables in population health monitoring systems used by different countries across different cultural contexts

– particularly to learn from the methods used to undertake a system “changeover” from self-report to wearables, or a change in the wearable device/wear location.

From the above list of knowledge gaps, Dr Bull summarized five emerging research areas:

- **Research area 1** – An *independent evaluation* of existing algorithms on MVPA, sedentary behaviour, and step-count measurement – this requires a data sharing and independent research group.
- **Research area 2** – More *developmental research* to improve both machine learning methods and other methods to classify and develop an MVPA metric; and to evaluate performance in older adults and more diverse and low- and middle-income populations.
- **Research area 3** – An *evaluation study* in multiple low- and middle-income countries to test and refine algorithms in populations (similar to GPAQ validation “9-country study”).
- **Research area 4** – A *demonstration study* to assess the utility and feasibility of WHO draft guidance on the use of wearable devices in population health surveillance systems across set of diverse cultures and contexts (for example by recruiting and supporting several WHO STEPS countries over next 12–24 months).
- **Research area 5** – Establish a *global mechanism and protocol* for ongoing independent evaluations of new algorithms – this requires establishing an independent “benchmark” dataset and testing to be undertaken independently from algorithm developers, i.e., by a central hub (this work could be linked to Research Area 1).

WHO notes that the above research tasks are neither sequential nor mutually exclusive. It is possible that more than one study could be conducted in one or more low- or middle-income country, and projects could commence concurrently if resources allow. Advancing the above lines of research would help address evidence gaps while building interest and more support for transition to wearables in population monitoring across more countries. Delegates suggested that an interim working group of researchers could convene to continue the discussion and advance the above actions.

The session closed with Q&A about the next steps, proposed research areas and more details on the experiences of the GPAQ multi-country validation study.

Closing Day 2: Commentary from an industry perspective: Discovery Vitality (Dr Mosima Mabunda)

Dr Mabunda provided comments from the perspective of Discovery Vitality, a behaviour-change platform that guides and incentivizes individuals toward healthier lifestyle choices and uses wearable-device data to assess physical activity behaviour. By incentivizing its members to adopt wearable devices and track activity, Discovery Vitality has built an extensive longitudinal dataset, drawing from both consumer wearables and mobile phones. The key points of interest from the meeting discussions and noted from their experience with the Vitality programme were:

- **Meeting people where they are:** Discovery Vitality leverages step counts as an accessible entry point, particularly for least-active individuals, to encourage the initiation of physical activity. For those without wearables, mobile phone data is used to ensure inclusivity and promote gradual habit formation.
- **Beyond steps:** In addition to steps, Discovery Vitality is exploring complementary metrics such as sedentary time and strength (e.g. handgrip strength) to provide a more holistic view of physical activity and health.
- **Rigorous use of device data:** Discovery Vitality integrates processed data from vetted consumer devices and mobile phones, ensuring reliability while minimizing the complexities and costs associated with using raw data.
- **Improving access to wearables:** Discovery Vitality expands access to wearable devices for its members through strategic partnerships and its global scale, helping to overcome affordability and accessibility barriers.
- **Philosophy of progress over perfection:** As a behaviour-change platform, Discovery Vitality ensures that the absence of a wearable device does not become a barrier to participation. Tiered incentives accommodate members who track steps via mobile phones, as well as other forms of physical activity, such as gym attendance.
- **Behaviour-change and surveillance:** Given the need to both monitor population-level activity and actively reduce physical inactivity between surveillance periods, it is important to consider what a “fit-for-purpose” device might look like. Such a device would not only enable accurate surveillance but also support behaviour change through user engagement and feedback.
- **Wearables for population surveillance:** In South Africa, Discovery Vitality’s work has demonstrated the acceptability of wearables. However, the need remains to demonstrate that algorithms used in surveillance are generalizable to local populations. This aligns with WHO’s broader concern around ensuring equity and contextual relevance in global health surveillance.

Day 3

The focus for Day 3 shifted to a discussion of the development of consensus in the reporting of research studies using wearable devices. It also included a discussion of the different processes required to rigorously establish a robust consensus.

Session 9: A framework for developing consensus (Paul Blazey)

Paul Blazey provided an overview of the consensus process that led to publication of the ACcurate CONsensus Reporting Document¹⁰ (ACCORD). ACCORD is the first reporting guideline and applicable to all consensus-based studies. Key points to consider while seeking to develop consensus include:

- **Resources/Tools** – For developing a reporting guideline, Moher et al. 2010 (19) provides the best available roadmap for developing reporting guidelines (an 18-step process).
- **Timeline** – Reporting guideline development takes longer than may be anticipated (1–3 years), depending on availability of resources and key partners in the process.
- **Clarity of purpose** – Establish a clear purpose and target audience for the consensus recommendations (in this instance, reporting standards). Failure to define the purpose may lead to poorer, shallower (less-useable) recommendations.
- **Expertise** – Involve a variety of experts and be authentic about involvement (i.e., if recommendations will apply to those outside the research-world, the process will need to include them).
- **Costs** – Think about the financial resources available, and a design consensus process that includes all key partners so that it can be both impactful and well received by those interested in the recommendations.

Other lessons from developing consensus for a reporting guideline were:

- Planning, including producing a protocol for consensus development and defining the focussed scope for the guideline will support the development of robust recommendations.
- Seek feedback from peers and those outside of your sphere to identify gaps in your consensus protocol.
- Plan regular (monthly) meetings between those leading the process to maintain momentum, especially where international collaboration is required.

¹⁰ www.ismpp.org/accord

- Be pragmatic – developing a reporting guideline that works is iterative and should include the development of an explanation and elaboration document to support effective use, followed by pilot testing.

The session finished by acknowledging several of the projects discussed (such as Session 4, Part 2), lend themselves to using consensus-based processes. An example included the need to develop clear operational *definitions* that can be applied across studies for the measures of interest, e.g. sedentary behaviour, steps, and MVPA.

Session 10: Report on ongoing process towards developing guidelines for reporting research on wearable devices (Dr Marco Giurgiu, Dr Jason Gill)

This session addressed Meeting objective 5. Dr Giurgiu described an ongoing global process being led by Dr Grace Dibben at the University of Glasgow to develop consensus-based, internationally endorsed reporting guidelines to improve the quality and transparency of studies measuring physical activity, sedentary behaviour or sleep via accelerometry. The ongoing work has included a scoping review and an online, interactive Delphi process. The review identified 47 existing publications containing reporting guidance for accelerometer-based assessment of physical activity, sedentary behaviour and sleep in health research and found that only 9% reported any consensus approach to their guidance development.

The review identified 124 unique items that were considered as necessary to report. However over 30 items were related to outcomes and metrics outside the scope of the reporting checklist and others were agreed to be too specialist (e.g. those related to reporting specific study designs such as 3D joint mapping) and should be consolidated. Following structured discussions within the international expert team, 51 items were selected for inclusion in the Delphi process. A consensus meeting and piloting exercise are planned to finalize the reporting guidance.

Dr Dibben, as lead for the Delphi study, has to date undertaken two rounds and received input from 111 contributors (round 1 held November–December 2024); and 93 contributors (round 2 – held March–April 2025). Of the 52 proposed items, there was consensus to report on 41 of them. Dr Giurgiu led an interactive rating of 25 items and invited meeting delegates to “vote” on their top 10 (see Fig. 2 for these items in order of rank).

Fig. 2. Draft results of a meeting vote to select the top 10 items for reporting guidelines for research using wearable devices



Discussion and Q&A highlights – Dr Marco Giurgiu

- Overall the delegates agreed there is a tension between a desire for requesting and reporting on a comprehensive set of information and balancing the perceived burden on researchers of doing so.
- Some items are aspects of the device specifications and as such may be redundant as these details can be found on websites and from the device – however others noted that any changes away from the “default” of the device should be identified.
- Delegates reached consensus on 8–10 of the items and suggested that reporting could be made easier if these were formatted as an author “checklist”; such as checklist could be published as an online appendix alongside the article. A checklist would reduce the burden of providing information and increase compliance.

- A lot of research uses consumer wearables for which details of some of the device and algorithm specifications are often not possible to obtain and/or are very dynamic and change unpredictably (and these changes may or may not be revealed by the device manufacturer). Nonetheless, it was noted that there is a lot of experience in the private sector (product developers/manufacturers) and that they could be involved in consultation. Also their knowledge of analysis of MVPA could be useful as many consumer devices also collected data on heart rate.
- Delegates discussed if the requirements for reporting details of the device and protocols might be different depending on the study purpose – for example, a validation study should be required to provide more details across more of the items compared to an intervention study which may be required to report fewer details as the focus is not on testing the device itself but testing the change in behaviour as a result of an intervention. Journals might consider trying to establish different reporting requirements for different study purposes or designs.
- It was also noted that there are often multiple publications from the same individual study, so once one publication has reported all the technical details required, it might be possible to simply cite the earlier publication for the readers to have access to all the device specifications.

Dr Gill, Deputy Editor of the British Journal of Sports Medicine (BJSM), provided perspectives on BJSM's interest in publishing articles related to physical activity and wearable devices. He highlighted what the journal views as desirable and suggested some areas for future publications. Key points included:

- BJSM asks “Is it new, is it true, does it change what we do?”
- BJSM is mostly interested in things that change practice, and in research from low-resource settings.
- BJSM generally publishes research related to health outcomes and therefore might not publish a validation study. However, if for example a study on a one-second epoch versus a 10-second epoch changes the relationship with health outcomes – this would likely be of interest to BJSM.
- BJSM is interested in state-of-the-art reviews – including on wearables – and on giving guidance to clinicians on what they should know about the use of wearables in research.
- BJSM noted it does have strong interest in publishing consensus statements on topics, such as reporting on wearable device research requirements, however BJSM may not formally mandate that submitting authors must follow the guidelines – this would be an issue for the Editorial Board and others to discuss.
- BJSM is very interested in publishing more editorials about the use of wearables and to the current state of the wearables field.
- Some areas of interest for future BJSM publications related to wearables and health outcomes might be how decisions on output metrics, placement, and device time can influence health outcomes, surveillance, insights into dose-response relationships; feedback to guide behaviour change and measuring changes in physical activity behaviours.

Discussion and Q&A highlights – Dr Jason Gill

The group discussion highlighted the different priorities of the BJSM and the Journal for the Measurement of Physical Behaviours (JMPB). Dr Giurgiu noted that JMPB has an upcoming special issue *“From heterogeneity to harmonization and/or standardization: addressing variability in accelerometer-assessed physical activity methods and outcomes”* (for which the submission deadline is December 2025).

Representatives from JMPB noted that the journal is considering developing and implementing a reporting guideline for device-based research and could make this a publication requirement. JMPB is particularly interested in and focused on developing guidelines for reporting on evaluation and measurement studies.

It was pointed out that consensus statements are occasionally co-published in different journals for broader impact.

The discussion ended by inviting follow up with the BJSM about any of the suggested areas of interest and the ongoing work of the JMPB and ISMPB. Details of a global, free ISMPB webinar with input from WHO and Dr Fiona Bull is being planned for 4–5 November 2025 (registration will be possible through ismpb.org).

Session 11: Closing plenary (Dr Fiona Bull)

WHO restated the direction of follow up actions after this meeting, as outlined in Session 8 on Day 2. This included drafting a meeting report that would be shared for review and revision and published on WHO website; and drafting the first version of the technical guidance document to circulate and conduct consultations with various experts and stakeholders. In addition, WHO will consider opportunities to convene a consultation with government representatives and relevant statistical agencies to learn from their protocols and experiences to inform the drafting of protocols and guidance. This additional meeting would aim to learn how countries have integrated wearable devices into national monitoring systems, and how they have mitigated any logistical challenges and cultural issues (as listed in Table A3.1).

Dr Bull summarized research areas emerging from discussions:

- Independent evaluation of existing algorithms on MVPA, sedentary behaviour, and step-count measurement;
- More developmental research to improve both machine learning methods and other methods to classify and develop an MVPA metric and to evaluate performance in older adults and more diverse and low- and middle-income populations;
- Evaluation studies in multiple low- and middle-income countries to test and refine algorithms in populations;
- Demonstration studies to field test the utility and feasibility of WHO draft guidance on the use of wearable devices in population health surveillance systems across set of diverse cultures and contexts.

WHO encouraged all delegates to consider opportunities for collaboration and potential funding sources for the research agenda and outlined plans to consult internally and with stakeholders about potential mechanisms and financial support to help establish research collaborations to advance this agenda.

Discussion and Q&A highlights

- Evaluation studies (Research areas 2 and 3) required data on more diverse populations, including older adults, and use of a traceable criterion measure for comparison; they would also require collaboration to pool available data to create a larger (labelled) training data set.
- To support WHO, a sub-group of meeting delegates voluntarily agreed to convene after the meeting to further discuss potential collaboration to address the knowledge gaps.
- Conducting a multi-country demonstration study (Research area 4) was agreed as important for building acceptance and more interest in use of wearables and to test the feasibility of implementing the (draft) WHO technical guidance and protocols; this study might require the loan of wearable devices.
- WHO agreed to explore with NCD surveillance colleagues at WHO HQ and regional offices the potential to scope interest in Member States who are planning to conduct WHO STEPS in 2026–2028. It was noted that a demonstration study could also involve other interested countries (not just WHO STEPS).
- It was noted that a large multi-country demonstration study would be logistically challenging and require long lead times to secure country engagement as well expert technical support to assist in training and data support.
- WHO noted the scope of the knowledge gaps and proposed research and the need to consider priorities and capacity within the research community, as well as necessary resource mobilization.

The final session ended with a return to the “big picture” and WHO invited delegates to consolidate collective thinking by rapidly mapping out the arguments for and evidence needed to support a response to the following three questions: Why change from self-report measurement of physical activity? Why use a wearable device? Why use a wrist worn device and the proposed protocols? Delegate responses were collected and summarized (Annex 3).

Closing

The meeting closed with extended thanks to and from the participants, the organizers, and WHO for everyone’s contributions, collaboration and constructive input to all sessions across three intense days, and for the additional work undertaken by the sub-group. WHO thanked in particular the Dr Karim Khan for co-convening and all the team for the preparation and logistical work to make the meeting successful.

References

1. Physical activity measurement and surveillance in adults: report of a scoping and planning meeting, 27–28 November 2023. Geneva: World Health Organization; 2023 (<https://iris.who.int/handle/10665/378157>, accessed 12 October 2025).
2. 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>, accessed 12 October 2025).
3. Guidelines on physical activity and sedentary behaviour. Geneva: World Health Organization; 2020 (<https://apps.who.int/iris/handle/10665/336656>, accessed 12 October 2025).
4. Global levels of physical inactivity in adults: off track for 2030. Geneva: World Health Organization; 2024 (<https://iris.who.int/handle/10665/378026>, accessed 12 October 2025).
5. 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>, accessed 12 October 2025).
6. Ross, R, Chaput, JP, Giangregorio LM, Janssen I, et al. Canadian 24-hour movement guidelines for adults aged 18–64 years and adults aged 65 years or older: an integration of physical activity, sedentary behaviour, and sleep. *Appl Physiol Nutr Metabol*, 2020;45 (suppl) : S57-S102 (<https://pubmed.ncbi.nlm.nih.gov/33054332/>, accessed 12 October 2025).
7. Brocklebank L, Steptoe A, Zhao Y, Bloomberg M et al. Comparing two wrist-worn accelerometers (Axivity AX3 and Matrix 003) for measuring movement behaviors in British and Chinese older adults. *J Measur Phys Behav*, 2025;8 (<https://doi.org/10.1123/jmpb.2025-0011>, accessed 12 October 2025).
8. Small SR, Chan S, Walmsley R, von Fritsch L et al. Self-Supervised machine learning to characterize step counts from wrist-worn accelerometers in the UK Biobank. *Med Sci Sports Exerc*. 2024;56(10):1945–1953 (<https://pmc.ncbi.nlm.nih.gov/articles/PMC11402590/>, accessed 12 October 2025).
9. Koffman L, Muschelli III J. Evaluating step counting algorithms on subsecond wrist-worn accelerometry: a comparison using publicly available data sets *J Measur Phys Behav* 2024;7 (<https://doi.org/10.1123/jmpb.2024-0008>, accessed 12 October 2025).
10. Zisou C, Taylor H, Lacey B, Hammami L et al. Reproducibility and associated regression dilution bias of accelerometer-derived physical activity and sleep in the UK Biobank. 2025 medRxiv 2025.01.16.25320679 (<https://doi.org/10.1101/2025.01.16.25320679>, accessed 12 October 2025).
11. Mikolaizak AS, Rochester L, Maetzler W, Sharrack B et al. Connecting real-world digital mobility assessment to clinical outcomes for regulatory and clinical endorsement—the Mobilise-D study protocol. *PLOS One* 2022 (<https://doi.org/10.1371/journal.pone.0269615>, accessed 12 October 2025).
12. Logacjov A, Bach K. Self-supervised learning with randomized cross-sensor masked reconstruction for human activity recognition. *Eng Appl Artif Intel*, 2024;128:107478 (<https://www.sciencedirect.com/science/article/pii/S0952197623016627>, accessed 12 October 2025).

13. Keadle SK, Lyden KA, Strath S J, Staudenmayer JW, Freedson P S. A framework to evaluate devices that assess physical behavior. *Exerc Sport Sci Rev* 2019; 47:206-214 (<https://doi.org/10.1249/jes.000000000000206>, accessed 12 October 2025).
14. Giurgiu M, von Haaren-Mack B, Fiedler J, Woll S, et al. The wearable landscape: Issues pertaining to the validation of the measurement of 24-h physical activity, sedentary, and sleep behavior assessment. *J Sport Health Sci* 2025;14:101006 (<https://doi.org/10.1016/j.jshs.2024.101006>, accessed 12 October 2025).
15. Keadle SK, Martinez J, Strath S J, Sirard J et al. Evaluation of within- and between-site agreement for direct observation of physical behavior across four research groups. *J Measur Phys Behav* 2023;6:176-184 (<https://doi.org/10.1123/jmpb.2022-0048>, accessed 12 October 2025).
16. Bull FC, Maslin TS, Armstrong T. Global physical activity questionnaire (GPAQ): nine country reliability and validity study. *J Phys Act Health*. 2009;6:790-804 (<https://pubmed.ncbi.nlm.nih.gov/20101923/>, accessed 12 October 2025).
17. Beauchamp MK, D'Amore C, Raina P, McIlroy WE et al. Establishing global standards on wearable technology for measuring mobility in aging populations: an international consensus exercise. 2025 (<http://dx.doi.org/10.2139/ssrn.5373737>, accessed 12 October 2025).
18. World report on ageing and health. Geneva: World Health Organization; 2015 (<http://apps.who.int/iris/bitstream/10665/186463/1/9789240694811%5Feng.pdf>, accessed 12 October 2025).
19. Moher D, Schulz KF, Simera I, Altman DG Guidance for Developers of Health Research Reporting Guidelines. *PLoS Med* 2010;7: e1000217 (<https://doi.org/10.1371/journal.pmed.1000217>, accessed 12 October 2025).

Annex 1: Provisional programme

DAY 1: 10 June – Tango A of Hotel Place d'Armes

Time	Session	Speaker(s)
08.00–09.00	<i>Breakfast and registration</i>	
09.00–10.00	Welcome, Land Acknowledgement, and Overview Brief introductions and outline of the meeting's focus on physical activity assessment and surveillance in adults. Discuss issues with using wearable technology for tracking physical activity, sedentary, and sleep behaviours in health monitoring. WHO Overview: • National physical activity monitoring tools and their global links. • WHO's work on wearable devices for population monitoring. • Key recommendations from previous WHO meetings.	Dr Fiona Bull and Dr Karim Khan
10.00–10.30	What are we measuring? Discussion on measuring physical activity in population health systems, with a focus on step counts per day as the main metric.	Dr Fiona Bull Commentators: Dr Rachel Colley and Dr Stephanie Prince Ware
10.30–10.45	<i>Break</i>	
10.45–12.30	What is this device going to do? What must it have, and what is nice to have? Discussion on defining the core features and functions of a "fit for purpose" wearable device for measuring physical activity and sedentary behaviours, suitable for various age groups.	Dr Fiona Bull, Dr Rick Troiano and Dr Matthew Ahmadi
12.30–13.30	<i>Lunch in Tango B of Hotel Place d'Armes</i>	
13.30–15.00	Validating, storing, analysing, and harmonizing the data Discussion on the availability of open-source algorithms for data processing, the need for developing new or modified open-source code, and the identification of knowledge gaps and necessary validation studies.	Moderator: Dr Fiona Bull Dr Kersten Bach, Dr Marco Giurgiu and Dr Aiden Doherty

Time	Session	Speaker(s)
15.00–15.30	<i>Break</i>	
15.30–16.45	Validating, storing, analysing, and harmonizing the data Session continued.	Moderator: Dr Fiona Bull Dr Kersten Bach, Dr Marco Giurgiu and Dr Aiden Doherty
16.45–17.00	Recap Day 1 Summary of discussion, emerging issues and recommendations from Day 1.	Dr Rick Troiano, Dr Mosima Mabunda and Dr Matthew Ahmadi

DAY 2: 11 June – Tango A of Hotel Place d'Armes

Time	Session	Speaker(s)
08.00–09.00	<i>Breakfast</i>	
09.00–09.15	Setting the scene for day 2 Context for Day 2 discussions.	Dr Fiona Bull
09.15–10.45	Other data requirements Discussion on what other data are (or may be needed for) supporting measurement and analysis of wearable data to report physical activity and sedentary behaviours.	Dr Fiona Bull
10.45–11.15	<i>Break</i>	
11.15–12.45	Parking lot Dedicated discussion time for any additional conversation topics that have emerged.	
12.45–14.00	<i>Lunch in Tango B of Hotel Place d'Armes</i>	
14.00–15.30	Age-related considerations and sedentary behaviours Discussion on whether solutions proposed are suitable for older people, seek input on life-course considerations, and address how sedentary behaviours fit into the overall conversation.	Speaker: Dr Marla Beauchamp
15.30–16.00	<i>Break</i>	
16.00–17.00	Parking lot Dedicated discussion time for any additional conversation topics that have emerged.	
17.00–17.15	Recap Day 2 Summary of discussion and emerging issues and recommendations from Day 2.	Dr Rick Troiano, Dr Mosima Mabunda and Dr Matthew Ahmadi

DAY 3: 12 June – Tango A of Hotel Place d'Armes

Time	Session	Speaker(s)
08.00–08.45	<i>Breakfast</i>	
08.45–09.00	Setting the scene for Day 3 Context for Day 3 discussions, and short introduction of ACCORD reporting guideline.	Dr Karim Khan and Paul Blazey
09.00–10.40	Scientific reporting guidelines for wearables Review current efforts aimed at strengthening the scientific reporting of research and development on wearable technologies in the fields of physical activity and sedentary behaviours.	Dr Jason Gill and Dr Marco Giurgiu
10.40–11.00	<i>Break</i>	
11.00–12.30	Plenary recap Provide meeting summary and outline recommended next steps and actions, including WHO's global consultation process moving forward.	Dr Fiona Bull
12.30–13.00	Closing session	Dr Fiona Bull and Dr Karim Khan

Annex 2: List of participants

External Academic Experts

Dr Matthew Ahmadi
Postdoctoral Research Fellow
Charles Perkins Centre
Faculty of Medicine and Health
University of Sydney
Australia

Professor Kerstin Bach
Research Director, NorwAI
Data and Artificial Intelligence Group
Department of Computer Science
Norwegian University of Science and Technology (NTNU)
Norway

Dr Julia K. Baumgart
University of British Columbia &
Norwegian University of Science and Technology
Norway

Dr Marla Beauchamp
Physical Therapist & Associate Professor
School of Rehabilitation Sciences
Associate Member, Department of Medicine
McMaster University
Canada

Paul Blazey
Research Manager
University of British Columbia
Canada

Dr Rachel Colley
Senior Research Analyst
Health Analysis Division
Statistics Canada
Canada

Dr Aiden Doherty
Wellcome Trust Senior Research Fellow
Professor of Biomedical Informatics
University of Oxford
United Kingdom

Professor Guy Faulkner
Professor and Chair, Applied Public Health
Population Physical Activity
School of Kinesiology
University of British Columbia
Canada

Dr Daniel Fuller
Associate Professor
Community Health and Epidemiology
University of Saskatchewan
Canada

Professor Jason Gill
Professor of Cardiovascular and Metabolic Health
University of Glasgow
University of British Columbia
Canada

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

Professor Jeffrey Hausdorff
The Igor Ornstein Chair for the Study of Geriatrics
Tel Aviv University
Israel

Dr Sarah Kozey Keadle
Associate Professor
Department of Kinesiology and Public Health
California Polytechnic State University
San Luis Obispo
USA

Dr Mosima Mabunda
Chief Clinical Officer
Discovery Health – Vitality
Johannesburg,
South Africa

Dr Charles E. Matthews
Senior Investigator
Metabolic Epidemiology Branch
National Institutes of Health
USA

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

Professor Emmanuel (Manos) Stamatakis
Professor of Physical Activity, Lifestyle and Population Health, Chair of ProPASS
School of Health Sciences
University of Sydney
Australia

Dr Olga Theou
Canada Research Chair in Physical Activity,
Mobility, and Healthy Aging
Dalhousie University
Canada

Dr Rick Troiano
Retired, United States Public Health Service
Surveillance of Physical Activity Expert
Maryland
USA

Canadian Institutes of Health Research (CIHR)

Marian Chong-Kit
Associate Director, Finance and Operations
Institute of Musculoskeletal Health and Arthritis
Canadian Institutes of Health Research
University of British Columbia
Canada

Professor Karim Khan
Scientific Director
Institute of Musculoskeletal Health and Arthritis
Canadian Institutes of Health Research
University of British Columbia
Canada

Dr Hetty Mulhall
Associate Scientific Director
Institute of Musculoskeletal Health and Arthritis
Canadian Institutes of Health Research
University of British Columbia
Canada

World Health Organization

Dr Fiona Bull
Head of Physical Activity Unit
Department of Health Promotion
Universal Health Coverage/Healthier Populations Division
World Health Organization
Switzerland

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

The following delegates reported interests were identified

- 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.
- Dr Prince Ware – reported her employer (PHAC) had an interest in results coming out of this meeting.
- Dr Mabunda – reported the cost of trip covered by employer Discovery Vitality.
- Dr Colley – reported her employer (StatsCan) had an interest in results coming out of this meeting.
- Dr Hausdorff – reported previous consulting work with Roche Pharmaceuticals.

None were deemed to represent a conflict of interest.

Annex 3: Discussion on factors that influence comparability of data on physical activity and sedentary behaviours from wearable devices

Four tables setting out relevant issues, defining key terms and outlining potential implications of each issue on data quality or comparability were prepared by WHO for the Physical Activity Measurement Meeting, Montreal, 10–12 June 2025.

These tables included a discussion question – along with, in some cases, a proposed position from WHO – regarding the use of wearables in population-based health monitoring systems. The tables were as follows:

- Table 1. Issues related to sample selection, device administration and compliance incentives
- Table 2. Issues related to wear protocols
- Table 3. Issues related to device properties
- Table 4. Issues related to processing protocols

Tables 1–3 were discussed both in the meeting and in post-meeting reviews. As explained in the main report, the issues in Table 1 were not discussed in the meeting (see write-up of Session 5). This annex therefore contains Tables 1–3, updated to contain details of, and outcomes and decisions on, the issues discussed in the meeting. These tables report final position reached by meeting delegates following discussion and post-meeting review, namely:

Table A3.1. Issues related to sample selection, device administration and compliance incentives

Table A3.2. Issues related to wear protocols

Table A3.3. Issues related to device properties

Table A3.4. Issues related to data processing protocols

Table A3.1. Issues related to sample selection, device administration and compliance incentives (note: these issues were not the primary focus of this third expert meeting and were not subject to detailed discussion)

1.1 Sampling (and subsampling) method	
1.2 Recruitment method	
1.3 Adjustment for sampling bias	
1.4 Device administration	These issues were not the focus of this meeting and not discussed in detail and therefore no recommendations.
1.5 Compliance strategies	
1.6 Collection of data related to physical activity	
1.7 Collection of socio-demographic data	
1.8 Collection of any environmental / geographic data	
1.9 Training for data collectors	

Table A3.2. Issues related to wear protocols

Category	Issue	Definition	Explanation of important and effect on estimate of physical activity	Summary of meeting recommendations on specifications
2. Wear protocols	2.1 Wear location	The wear location (position) of the device on the body	Different parts of the body move to different degrees during different activities, and this influences the amount of movement detected. This may lead to important discrepancies in detection and analysis of certain physical activity activities and sedentary behaviours. Although, in general, estimates from different wear locations (and devices) are highly correlated, they produce different absolute values of physical activity.	Minimum: Wrist worn on non-dominant hand <i>Rationale</i> Most practical and acceptable device for use in population monitoring systems. AND Recommend recording the wear location for each responder (as some may decline non-dominant). <i>Notes</i> • Agreed this wear location was suitable for older adults. • Considered to be suitable for some people living with disability, including wheelchair users.
	2.2 Wear time	Instructions to participants on the number of days, hours per day, and for specific activities (e.g. sleep, water-based activities)	Different instructions and inclusion criteria for wear time (days/hours) will impact physical activity estimates. Notable issues include: • More measurement days provide more stable measurements by minimizing the influence of atypical days. • Participants may increase physical activity (e.g. social desirability) in response to participation and wearing a device. • Physical activity has been shown to vary between weekend and weekdays. • Most devices are now waterproof, avoiding the need to remove for washing/aquatic activities.	Wear days:* Minimum 7 days Optional: collect >7 days <i>Rationale</i> • For implementation now, a 7-day protocol is consistent with both usual practice and global physical activity guidelines over past 20 years. • With more national data collection from country experiences, it will be possible to review this protocol and update/adapt. • Given 7 days there is no need to specify wear days requiring weekend day(s).

Category	Issue	Definition	Explanation of important and effect on estimate of physical activity	Summary of meeting recommendations on specifications
				Wear time: * Not decided Discussion within the meeting and post-meeting comments suggested use of a minimum of 24-hours; however, some other options were suggested e.g. a minimum of 10-, 18- or 24- hours. As the duration of a waking day can vary substantially between individuals (e.g. 10 to 20 hours) the criteria for a valid day could be based on a specified proportion (%) of each wake time/day – for example 90% of waking day hours with valid wear time (e.g. 8.1 hours for a 9-hour waking day and 14.4 hours for a 16-hour waking day). *See Table 3.2 row 4.5 for criteria on “valid days/time” for data analysis.
	2.3 Device orientation	The orientation of the device (specifically, its axes of measurement) relative to the body	There are no concerns with the subject correctly wearing (orientation) a wrist worn device.	No other issues identified re: wrist worn-device orientation. <i>Note</i> This assumes that there are no removable parts to the device; that there is a visible indication on the device of which is the correct way “up”; and that the analysis method can detect and analyse data collected if the device was inadvertently worn facing the wrong way (“down”).

Table A3.3. Issues related to device properties (overarching recognition that decisions on these issues have implications on battery life)

Category	Issue	Definition	Explanation of important and effect on estimate of physical activity	Summary of meeting recommendations on specifications
3. Device properties including sensor specifications	3.1 Number of axes	The number of axes (between 1 and 3) on which the device measures acceleration	A single axis has less sensitivity to movement compared to three axes – most devices now use triaxial sensors.	Minimum: Triaxial
	3.2 Dynamic range	The range of acceleration that can be measured	The maximum measurable acceleration without clipping. The range set must be able to capture the majority of human movement which occurs between a range of 1–2 g (as example for Actigraph GT9x it is +/- 8g).	Minimum: +/- 6 g <i>Rationale</i> - +/- 6 g is considered to be an adequate buffer given most human movement is between 1–2 g and at 12 bits the resolution (i.e. sensitivity) has a Least Significant Bit (LSB) of 0.003 g (that is 12g/4096) and this could be considered “good enough” for physical activity recognition.
	3.3 Resolution	The smallest change in acceleration that the device can detect	The resolution is determined by the number of bits used in the analogue-to-digital conversion (ADC) process.	Minimum: 12 bits <i>Rationale</i> Selected to balance competing demands, noting that higher resolution will require more power and that more bits require greater storage capacity.
	3.4 Sensitivity	The output of the accelerometer per unit of acceleration (typically measured in g, where 1g = 9.81m/s ²)	Higher sensitivity indicates that the sensor can detect smaller changes in acceleration. Manufacturers typically state the specification and conduct laboratory testing. Independent calibration testing can be conducted to ensure output corresponds to known acceleration level.	As sensitivity is a function of the dynamic range and resolution – using proposed +/-6 g and with 12 bits – the resolution (sensitivity) is at a Least Significant Bit (LSB) of ~0.003.

Category	Issue	Definition	Explanation of important and effect on estimate of physical activity	Summary of meeting recommendations on specifications
	3.5 Noise floor	Level of background noise that the sensor produces – measured in standard deviation of readings while sedentary	Lower level of noise floor indicates better resolution as it allows for detecting smaller variations in acceleration without interference; this can be important for assessing physical activity among older and younger populations where their movement patterns are different. Noise floor data may be hard to obtain as it is proprietary information	Minimum: 1 milli-g to 5 milli-g <i>Rationale</i> Most accelerometers are below 5 milli-g and it is achievable even for cheap (US\$ 5) accelerometers.
	3.6 Temperature sensitivity	The temperature range in which the accelerometer will operate without failure	Device is required to be robust for use in all environmental conditions across diverse countries. Temperature sensitivity can affect sensitivity and resolution, and thus the performance of the device in different conditions (this is a critical issue for comparisons between countries). A stable performance across a range of temperatures is desirable to demonstrate high reliability. Available (“cheap”) accelerometers appear to be rated for the -40°C to 85°C range (see ADXL345 Datasheet and Product Info Analog Devices)	Minimum: operational range should be minus 20°C to plus 50°C <i>Notes</i> Within the range of the manufacturer’s temperature recommendations, signal drift will occur at lower /higher temperatures. This is usually not a problem as sensor calibration can correct for the drift. However, it does put “stress” on the device which can accelerate device degradation/quality, resulting in short product life cycle and the need to purchase more devices sooner.

Category	Issue	Definition	Explanation of important and effect on estimate of physical activity	Summary of meeting recommendations on specifications
	3.7 Sampling frequency	The sampling frequency of the acceleration signal per second	Higher sampling frequency will require larger data storage (memory) capacity, while lower frequency can reduce requirements for data storage. Higher sampling frequencies can result in higher estimates of physical activity as more movement is detected. Sampling frequency can interact with decisions on filtering frequency (see Table 4) with some combinations being less effective at filtering out noise.	Minimum: 30Hz Optional – 40Hz or higher <i>Rationale</i> For human movement from the wrist, 30 Hz was recommended to be “good enough” considering majority of human movement occurs under 5 Hz. For example, a 30 Hz sampling rate allows capture of frequencies up to 15 Hz and safely above 5 Hz and would help extend battery life. The addition of an option to use 40 Hz or higher was proposed because vigorous human movements are at 16–18 Hz.
	3.8 Memory capacity	How much data can be stored in raw form	Memory capacity must be sufficient for the minimal specifications set for the device and data storage must be robust to avoid data loss – particularly if the battery fails.	Minimum: Minimum memory capacity must be sufficient to support data collection at the specified number of days, hours per day and sample frequency. AND Raw accelerometer data must be stored in a non-volatile way to avoid data loss should battery life end.
	3.9 Battery life	Capacity to run device – over days/ weeks	Specification is dependent on other device characteristics (see above) and must cover at least the minimum number of days and time required for transport and return – preferably with no user recharging required.	Minimum battery needs to be sufficient to support the required data collection days and hours per day at specified sample frequency. AND Preferably the device should not require the user to re-charge during the data-collection period. AND If the device does require re-charging there must be mechanism(s) to protect the data and avoid user access or user download.

Category	Issue	Definition	Explanation of important and effect on estimate of physical activity	Summary of meeting recommendations on specifications
	ISSUE ADDED IN MEETING 3.10 Battery recharging and replacement	Battery recharging	Ability to recharge the battery is required to enable re-use of the device by multiple persons within a single data-collection period and to support re-use over years in subsequent data-collection periods. Battery recharging can be time-consuming so an easy and efficient process is required to minimize the burden during rotation within a data collection period. The method of recharging and battery replacement can be logistically complex and costly if it requires a return to the manufacturer.	Minimum: Device must be re-usable and re-chargeable by central coordinator to enable re-use within the surveillance programme. AND Devices must be re-chargeable at scale (i.e. with use of a docking station for large batch recharging).
	3.11 Feedback to user	Whether the device provides any visual or audio feedback to the user – such as on their level of physical activity in step counts or minutes	User feedback is an effective mechanism for behaviour change and can potentially increase compliance with wear protocols. However, within the context of population surveillance this feature could be a negative factor as user feedback may influence or alter the usual pattern of physical activity behaviour. This could potentially produce higher physical activity/lower sedentary time estimates. Closed devices provide no feedback on physical activity but may be perceived as less “useful” and affect user adherence (wear days/time). Providing feedback (such as time) could increase wear compliance.	<u>Provision of feedback on physical activity/sedentary</u> No feedback should be provided to user on physical activity (steps, minutes, or other metrics) or on sedentary behaviours/time. <u>Provision of other feedback or device uses</u> Not decided Some discussion but no clear decision made in meeting or post-meeting review on including a clockface which may provide some user utility but also add to cost of device and impact on battery requirements and potential other factors.

Category	Issue	Definition	Explanation of important and effect on estimate of physical activity	Summary of meeting recommendations on specifications
	3.12 Use of multi-sensor approach	Additional sensors can provide useful data to support and improve analysis	Use of additional sensors can provide diagnostic feedback on the performance of the device; or provide additional data on body position, physiological responses or movement. The data from additional sensors can be used to improve classification of accelerometer data and thereby influence the reported output metrics. Examples of other sensors that are in use in wearable devices include: gyroscopes, photoplethysmography (PPG); ambient temperature; skin conductance sensor; skin temperature; global positioning systems (GPS).	Additional sensors required for diagnostic purposes Minimum: (1) Battery health – useful for diagnostic purposes (i.e. when device has problems) to provide insight on cause and assist with decisions on data inclusion. (2) Temperature – useful for (i) diagnostic purposes; (ii) providing “near surface” temperature to assess near-skin temperature may also be useful to assess wear versus non-wear periods; Optional: Device allows temperature and battery health sensors (described above) to be switched on and off, and therefore used only in the country context as required – this could be considered if it did not detrimentally add to cost. OTHER SENSORS (to assist analysis /other purposes/ research) (1) Heart rate (via PPG) Not decided <i>Note</i> • This additional data can help improve the classification of MVPA but has many negative trade-offs, including additional requirements of data storage and battery life, as well as needing individual calibration • Noted that heart rate can be affected by medication. • Noted the issue of PPG granularity / accuracy versus ECG capabilities (and issue of arm movements and different activity types) Overall, no final decision was reached due to insufficient time for a full discussion on this issue.

Category	Issue	Definition	Explanation of important and effect on estimate of physical activity	Summary of meeting recommendations on specifications
				(2) Global positioning systems (GPS) Minimum: No <i>Rationale</i> • Due to increased demands on the device battery and memory capacity, as well as privacy issues that may influence response/compliance, • Due to high complexity of data analysis (3) Gyroscope Minimum: No Optional: Yes <i>Rationale</i> • Adding a gyroscope has significant impact on battery requirements as it adds a lot of data while potentially not adding extra value when included in a wrist worn device. It was noted that it may be useful for assessing people with functional limitations.
	ISSUE ADDED IN MEETING 3.13 Waterproof and dustproof and intended for prolonged skin contact	How to waterproof and dustproof the devices	Devices must be waterproof – not only water resistant – to enable monitoring of water-based physical activity and to be secure for use in all weather conditions and whilst undertaking self-care (such as baths or showers). Devices must be dustproof to avoid device failure. Materials must be suitable for use with prolonged skin contact to allow for 7 days wear time. Both waterproof and dustproof specifications have IP (Ingress Protection) ratings.	Minimum: (1) Water resistant (IPx8) to allow for capture of swimming and other water-based activities. Minimum: (2) Dust tight (IP6x) to address issues that may be common in some country contexts. Minimum: (3) Use materials certified as hypoallergenic and biocompatible and intended for prolonged skin contact to allow for skin sensitivity (as there is no specific IP rating for skin sensitivity, certifications such as like ISO 10993 may be suitable).

Category	Issue	Definition	Explanation of important and effect on estimate of physical activity	Summary of meeting recommendations on specifications
	ISSUE ADDED IN MEETING 3.14 Device attachment	The wearable fitting for the device (e.g. a wrist band)	The device attachment must be secure, easy to use, and convenient mechanism (such as a strap); adaptable to fit the typical range of wrist sizes (in adults); and removable to allow for cleaning or replacement if damaged – these factors may affect compliance and thus wear time and data.	Minimum: Wrist band must be adjustable, removable and cleanable or disposable (and replacement units available).
	ISSUE ADDED IN MEETING 3.15 Data handling and transfer	How data are retrieved and transferred from the wearable device and handled to create a secure dataset ready for processing	Data handling can be complex and time-consuming; can require robust wi-fi access and external hard drives; and may require training of staff in specialist tasks. Data handling can affect the overall feasibility and efficiency of how a wearable device is integrated into the protocols of a surveillance system and therefore can be a factor in decision making.	Not decided <i>Notes</i> Insufficient discussion on requirement and specifications related to data download and transfer (e.g. USB port versus bluetooth versus wired transfer).

Table A3.4. Issues related to data processing protocols

Category	Issue	Definition	Explanation of important and effect on estimate of physical activity	Summary of meeting recommendations on specifications
4. Processing protocols	4.1 Access to raw signal	Whether the raw acceleration signal can be extracted from the device versus access only to summary-level statistics	Access to the raw signal provides maximum control over data-processing and analysis protocols. Raw data enable retrospective processing protocols as knowledge and analytic techniques advance. However, storing raw data impacts memory and battery requirements.	Minimum: Raw accelerometer data must be accessible
	4.2 Frequency filtering	The frequency filters applied to the signal that remove noise (and in some cases, the gravity component)	The choice of frequency for filtering noise can impact the estimates of physical activity because signals above or below the selected threshold are considered as noise and removed. The impact of filter selection is greatest on estimates of physical activity time spent at very low or very high intensity levels. The frequency filter is related to which metrics are required (and some metrics remove the component of gravity in different ways).	Minimum: 20Hz Butterworth filter <i>Notes</i> • This may or may not be generalizable to all current and future algorithms and other specifications may be required as the approach could vary for individual algorithms (i.e. different methods used to account for noise in different ways). • It was agreed that details of the handling of this issue should be clearly reported in the communication of analysis and output metrics on physical activity. • The method used should be based on ported and supported by independent evaluation using a benchmark dataset.
	4.3 Epoch length	The time period over which the signal is summarized	Usually set at 60 seconds but can be shorter. Longer epochs generally lead to lower estimates of physical activity as the spikes of short intense movements are lost when averaged.	Maximum: 60 seconds Optional: 10 or 30 seconds (not decided) <i>Notes</i> A post meeting suggestion was an optional protocol that allowed use of a lower epoch (such as 10 or 30 seconds) for adult populations. However use of different epochs could detrimentally affect the comparability of population level estimates between countries. This issue was not finalized.

Category	Issue	Definition	Explanation of important and effect on estimate of physical activity	Summary of meeting recommendations on specifications
	4.4 Minimum bout length and interruption tolerance	(1) The minimum length of time (i.e. "bout") an individual needs to spend at a particular intensity level (e.g. light, moderate, vigorous) (2) Interruption tolerance refers to how long a different intensity level may continue before the signal and time are counted in the analysis	Higher minimum bout lengths can result in less time recorded for a specific intensity category, whilst conversely shorter (or no) minimum bout length specification will increase the inclusion of signals and thus increase time counted. Higher tolerance for interruptions will reduce the estimate of time spent in any specific intensity category. Note – there is an interaction with epoch length, as longer epoch lengths reduce the risk of interruptions in intensity levels. Longer epochs can increase the estimate of time spent in an intensity level when a minimum bout length is applied.	Not recommended to use of minimum length of time (i.e. defining a "bout") in analysis Rationale To be consistent with 2020 WHO guidelines, in which bout length was removed. Tolerance for interruptions Not decided <i>Notes</i> The discussion of continuous time and tolerance for "interruption" was not fully resolved. It was noted that this will have some importance, especially if the methods of analysis uses cut point approach.
	4.5 Identifying non-wear	How periods of very low activity are distinguished from those during which the device has not been worn	Different methods of detecting non-wear time can lead to different amounts of time either being excluded from the analysis or counted as very low/light intensity activity. This is linked to wear time (2.2) as greater wear time can reduce the impact of different non-wear time methods on the physical activity estimates.	Minimum: Requires consecutive stationary episodes lasting for at least 60 minutes in which all three axes had a standard deviation < 13.0 milli-g <i>Notes</i> Post-meeting review raised the issue that using a 60-minute criterion may result in false negatives (e.g. if an adult takes off the device for an organized sport or swimming – what might have been MVPA would now be classified as sedentary). If this is considered to be infrequent or rare then it would not be issue. If not, 30 minutes could be used as a balance of accuracy and granularity.

Category	Issue	Definition	Explanation of important and effect on estimate of physical activity	Summary of meeting recommendations on specifications
Output metrics	4.6 Minimum-wear time inclusion criteria	The minimum wear time that an individual must have completed to be included in the analysis – this comprises the valid number of wear days with valid wear time per day	Lower wear time criteria (i.e., fewer days and fewer hours per day) will mean more data from individuals are included in the analyses – this can potentially improve sampling representativeness (both more people included and more data from more people). However, there is a risk that those who did not wear the device for very long contribute more to the final dataset and add bias to the data.	No decision Suggestions and comments from meeting and post-meeting review were: Minimum of 3 days with each day having valid hours OR Minimum 1 day with valid hours AND Where valid day requires data from: each wear hour period of the 24-hour cycle OR from at least 8hrs or at least 10hrs <i>Notes</i> • Evidence exists to show that restricting data to the number of days may lead to bias in surveillance estimates (see NHANES) this supports setting a 1-day minimum. • Setting a low threshold of 8 hours when wear time requires 24 hours will bias estimates of sedentary behaviour and possibly other metrics (e.g. steps). Setting minimum of 8–10 hours may be acceptable if stated as wake time only.
	4.7 Classification of activity intensity	The method used to summarize from the raw data the magnitude of the acceleration for each epoch Methods include Euclidean Norm Minus One, Mean Amplitude Deviation, total counts of activity	The different metrics of summarizing activity intensity are mostly highly correlated (but not identical) in terms of magnitude. This issue is linked to frequency filtering (4.2), as some methods account for gravity in their calculations, while others need frequency filters to be applied.	No decision <i>Notes</i> • Different analysis methods have different approaches. • The method selected by countries for their data analysis should be transparently reported.

Category	Issue	Definition	Explanation of important and effect on estimate of physical activity	Summary of meeting recommendations on specifications
	4.8 Classification of physical activity intensity	The method used to define and analyse each epoch of the raw signal to classify as different physical activity intensities (e.g. light, moderate, vigorous)	Two main methods are available: (i) Cut points (or threshold) can be used to distinguish one intensity level from another. (ii) Machine learning methods use a combination of the signal intensity an pattern to assign an intensity and behaviour to each epoch. The choice of cut points has an impact on the estimates of physical activity, e.g. the wider the distance between the cut-points, the more time that will be calculated as being in that intensity category.	No decision <i>Notes</i> The methods of machine learning algorithms (such as OxWearables, ActiPASS) should publish results from an independent evaluation conducted with an open benchmarked dataset. AND Should have published results demonstrating accuracy and precision in Phase III verification studies as described by Keadle et al (2019).
	4.9 Classification of specific activities e.g. step counts	The method used to detect/define specific activities, such as step counts, from the raw signal	Different methods have different accuracy	As above
	4.10 Classification of specific activities, e.g. sedentary	The method used to detect/define specific activities, such as sedentary, from the raw signal	Different methods have different accuracy	As above

Annex 4: Reasons for the transition from self-report to wearables

The Day 3 final Session 11 asked delegates to suggest points to include when answering the following key questions on transitioning from self-report to wearables as a means to measure physical activity and sedentary behaviour.

Why are we changing from self-report?

- To improve data systems and health surveillance on physical activity.
- To have better measures to quantify the health burden of inactivity.
- To get a measure closer to the truth that more accurately reflects “real-world” behaviour.
- To enable better harmonization of measurements across countries and across time.

Why wearables?

- Wearables provide more precise measurement of free-living physical activity compared to self-report instruments.
- There is evidence that wearables have been tried and tested in scientific epidemiological studies and show similar (and stronger) dose response associations with various health outcomes
- There is evidence that presents and favourably compares the magnitude of “error” for wearables compared to self-report instruments.
- Avoids self-report bias.
- Avoids recall problems.
- Self-report cannot measure certain aspects very well (i.e., sedentary behaviour, step counts, light-intensity physical activity, short intermittent movements).
- Provides greater consistency of measures.
- Take advantage of current technologies and are “innovative” / new / cutting edge.
- The outputs (such as step counts) can help better communicate on physical activity with the community (and with professionals).
- Can address cultural differences that may affect self-report data due to interpretation of questions.

- Will be more inclusive (remove language barriers).
- Will be more sensitive to detect change.
- Can provide data that will better inform developing interventions and target populations (such as step counts).
- Will align with the new/emerging data from epidemiological studies, which is increasingly using wearables (and shows a stronger dose-response with many health outcomes).
- Will align surveillance measurement with tools and metrics likely to underpin the physical activity guidelines in 2030 and beyond.
- Will help to better identify which groups need more focus to enable more physical activity.

Why a wrist worn device and these protocols?

The group helped list convincing evidence, including:

- Evidence that a wrist-worn device is comparable to other measures of exposure for measuring steps, sedentary and MVPA, i.e., devices worn in a different location and self-report.
- Future evidence will likely be from studies using the wrist location because:
 - wrist-worn compliance is better;
 - feasibility and acceptance by user;
 - wrist-worn devices offer potential to collect additional health-related data if desired (e.g. heart rate).
- Show evidence on the acceptability and feasibility in diverse contexts and populations, including from low- and middle-income countries (it was noted the evidence is currently not robust).
- Large numbers of people are now wearing monitoring devices and that the most popular are wrist worn – so this is an increasingly familiar gadget and aligns with what people are already seeing and using.
- Wrist-worn devices are suitable for older adults and people with different movement patterns/functions if data are properly processed.
- Wrist worn devices did not work well in the past, but we have new technological advances and machine learning methods to analyse data and these limitations have been addressed (with the exception of MVPA – while many regard wrist-worn devices as acceptable to measure MVPA, there is agreement on the need for more work).

Other issues that would also need to be addressed in guidance or “proposals” to a country include:

- Why not use data from commercial devices?
- What about data security?
- Cost implications for both the device and any replacements; and administration costs such as additional data collection time and travel to collect devices.
- Whether user feedback is provided and if not, why and if yes, what type, when and how.
- Will this have user acceptability and trust (in government, use of data, the device itself)?
- Choice of device and comparability between devices (across brands and models)?

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