

# A Consumer Cohort Yields Accurate Step Estimates in Young and Middle Aged Adults

**Abram Schönfeldt**

Google Research; Nuffield Department of Population Health, University of Oxford

**Katarina Vukosavljević**

Google Research; Institute of Biomedical Engineering, University of Oxford

**Justin Phillips**

Google Research

**YongSuk Cho**

Google Research

**Robert Harle**

Google Research; Department of Computer Science and Technology, University of Cambridge

**Shwetak Patel**

Google Research; Paul G. Allen School of Computer Science and Engineering, University of Washington

**Aiden Doherty**

Nuffield Department of Population Health, University of Oxford

**Daniel Roggen**

[roggen@google.com](mailto:roggen@google.com)

Google Research; School of Engineering & Informatics, University of Sussex

---

## Article

### Keywords:

**Posted Date:** September 4th, 2026

**DOI:** <https://doi.org/10.21203/rs.3.rs-10891779/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Additional Declarations:** **Yes** there is potential Competing Interest. This study was funded by Google. DR, JP, RH, YC, SP are Google employees and may own stock as part of the standard compensation package. AS and KV worked as student researchers at Google while this research was conducted.

---

# A Consumer Cohort Yields Accurate Step Estimates in Young and Middle Aged Adults

Abram Schönfeldt<sup>1,2</sup>, Katarina Vukosavljević<sup>1,3</sup>, Justin Phillips<sup>1</sup>, YongSuk Cho<sup>1</sup>, Robert Harle<sup>1,4</sup>, Shwetak Patel<sup>1,5</sup>, Aiden Doherty<sup>2,\*</sup> and Daniel Roggen<sup>1,6,\*</sup>

<sup>1</sup>Google Research, <sup>2</sup>Nuffield Department of Population Health, University of Oxford, <sup>3</sup>Institute of Biomedical Engineering, University of Oxford, <sup>4</sup>Department of Computer Science and Technology, University of Cambridge, <sup>5</sup>Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, USA, <sup>6</sup>School of Engineering & Informatics, University of Sussex, \*Joint senior authors

Widespread adoption of consumer wearables offers large-scale, continuous physical activity data, but concerns around non-representative device adoption and device differences have precluded their use in population health surveillance. We examined trends in mean daily steps from >200,000 Fitbit participants in Japan (2015–2023), and >300,000 in the United Kingdom (2021–2023) and >7 million in 28 countries globally (2015–2022) to national estimates. Fitbit step estimates showed close agreement with national estimates in younger and middle-aged adults in Japan (20–50 years: Mean Percentage Difference = -0.3%), and middle-aged adults in England (50–69 years: MPD = 0.1–2.0%), though were higher than national estimates in older adults in both countries (Japan 70+ years: MPD = 15.9%, England 80+ years: MPD = 20.1%). Changes in Fitbit steps showed a moderate-to-strong linear relationship with changes in the prevalence of people who report meeting physical activity guidelines (2010–2022) across 28 countries with complete demographic coverage ( $R^2 = 0.47$ – $0.50$ ). Consumer wearable data provide indications of long-term population health trends and could complement national physical activity surveillance in young and middle-aged adults in countries with sufficient demographic coverage, which tend to be in the Global North.

## 1. Introduction

Reliable and timely estimates of population physical activity are needed to monitor public health, develop and evaluate policies, and identify changes in health-related behaviour (Bull et al., 2020; Mason et al., 2022). However, most national surveillance systems rely on self-reported physical activity, which is susceptible to recall and social desirability biases (Sallis and Saelens, 2000; Shephard, 2003). Device-based measurements provide more objective estimates, but have only been incorporated into national surveillance in a small number of countries (de Wolf et al., 2024). One notable exception is Japan, which has measured steps using pedometers since 1995 as part of the National Health and Nutrition Survey (NHNS-J) (Nakagata and Ono, 2024).

Commercial wearable platforms continuously collect physical activity data from millions of users. These data are available at a scale and frequency that conventional national surveys are

currently unable to achieve, and could therefore complement existing population health surveillance. More broadly, they provide an opportunity to test whether this source of passively collected data can generate useful population health insights. However, commercial wearable users are typically younger, more active, and more affluent than the wider population (Mair et al., 2022; Omura et al., 2017; Strain et al., 2022). Differences in participant characteristics and measurement devices may therefore limit the extent to which estimates from commercial cohorts reflect national physical activity patterns.

It therefore remains unclear whether sample bias limitations preclude the use of commercial wearable data for population surveillance, or whether their large-scale and temporal resolution can nevertheless provide reliable information about national trends. To our knowledge, no large-scale longitudinal study has evaluated commercial wearable estimates against concurrent measurements from nationally representa-

tive surveillance programmes. This information is needed to assess whether commercial wearable data can augment public health surveillance systems.

Here, we compared mean daily step counts from 239,296 Fitbit participants with nationally representative estimates from the Japanese National Health and Nutrition Survey (NHNS-J) between 2015 and 2023 (Nakagata and Ono, 2024). To assess validity outside of Japan, we compared mean daily steps in 321,287 older Fitbit users with representative estimates from the English Longitudinal Study of Ageing between 2021 and 2023. Finally, we compared trends in mean daily Fitbit-measured steps with trends in self-reported insufficient physical activity over 2015 to 2022 in 28 countries. Our data from over 7 million Fitbit participants show the potential of commercial wearable data to inform population health surveillance.

## 2. Results

### 2.1. Demographic characteristics of Fitbit users

We compared the demographic distribution of Fitbit participants, who provided consent for scientific data sharing, in Japan, the United Kingdom, and 28 other countries with at least 5 years of estimates across all combinations of age and gender (Supplementary Table 3). The number of consenting Fitbit participants in countries meeting selection criteria increased over 20-fold from 262,475 participants in 2015 to 7,371,717 participants in 2025 (Figure 1). In Japan, there was an increase from 1,466 Fitbit participants in 2015 to 239,296 participants in 2023.

Overall, the Fitbit cohort – restricted to users who consented to data sharing and met wear-time criteria – tended to be younger with a higher proportion of women. In Japan, adults aged over 70 comprised fewer than 7% of the Fitbit cohort compared to 28% of the Japanese population. In contrast to the global distribution, women represented only 20% of participants in Japan in 2015, increasing to 40% in 2023. In the United Kingdom, women were overrepresented in the 50-59

(+17.4%) and 60-69 (+6.9%) age brackets and under-represented in the older age ranges, whilst men were under-represented in all groups, and particularly in the older age brackets.

### 2.2. Differences in step counts in Japan and the UK

We first compared step counts from the wrist-worn Fitbit cohort in Japan to the hip-measured, population-level estimates from the annual National Health and Nutrition Survey (NHNS-J) (Nakagata and Ono, 2024). Unadjusted Fitbit step estimates were substantially larger than NHNS-J estimates. The Mean Absolute Percentage Error (MAPE) between estimates across demographic groups was 52.8% (95% CI: 49.7–56.0%), or 3,185.7 steps/day (95% CI: 2,969–3,391) expressed as the Mean Absolute Error (MAE), with Fitbit users having higher step counts (Supplementary Table 1). Adjusting for differences between the wrist-worn Fitbit and hip-worn NHNS-J measurement devices using the calibration equation from Nakagata et al. (2022) had a major impact, and reduced the MAPE to 11.4% (95% CI: 9.0–14.7). That calibration equation was developed with participants aged 21–50. Focusing on differences within this age-range, the signed Mean Percentage Difference (MPD) decreased from 7.2% (95% CI: 3.0–11.8) over the 20–70+ age range to -0.3% (95% CI: -5.4–3.4) over the 20–50 age range. The discrepancy between the MPD (-0.3%) and the MAPE (6.7%, 95% CI: 5.3–8.1) over the 20–50 age range indicates that, although the overall difference is close to zero, non-zero differences persist within demographic sub-groups within this age-range.

To assess agreement when consumer and national studies both use wrist-worn accelerometers, we next compared Fitbit-measured steps in the United Kingdom with data from the 2021–2023 wave of the English Longitudinal Study of Ageing (ELSA) (Brocklebank et al., 2026). ELSA provided marginal mean step estimates by 10-year age groups (50–80+ age range) and, separately, by gender. The MAPE over the age marginals in England was 5.8% (95% CI: 4.3–8.3), or 376 (95% CI: 289–551) steps expressed as the MAE (Supplementary Table 2). Computing analogous

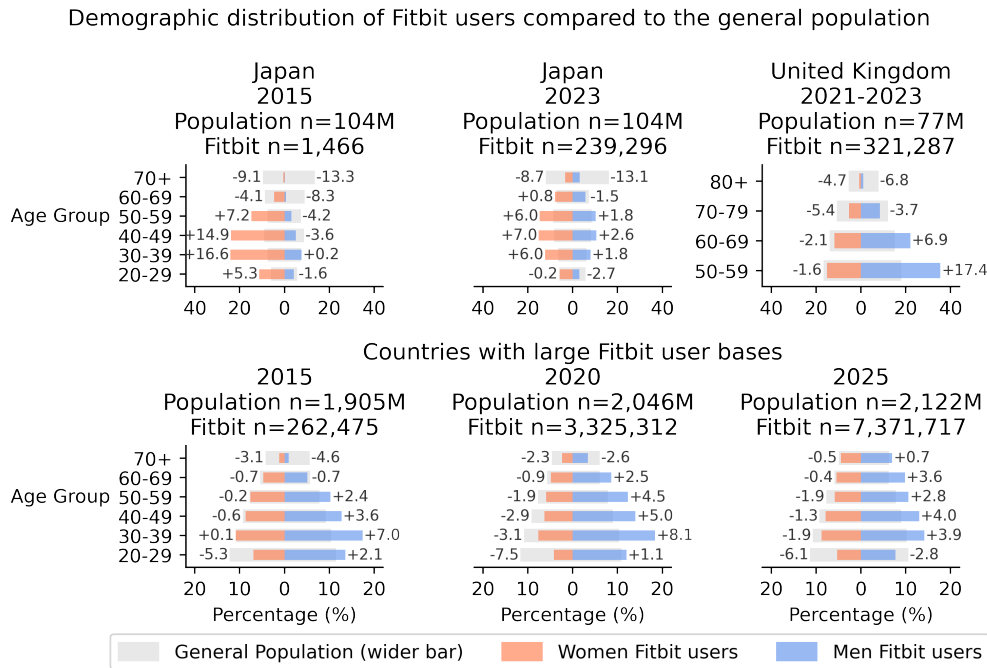


Figure 1 | Demographics of Fitbit users in Japan, the UK, and 28 other countries versus national estimates from [United Nations, Department of Economic and Social Affairs, Population Division \(2024\)](#). Selected countries contain at least 5 years of Fitbit estimates across all combinations of age and gender.

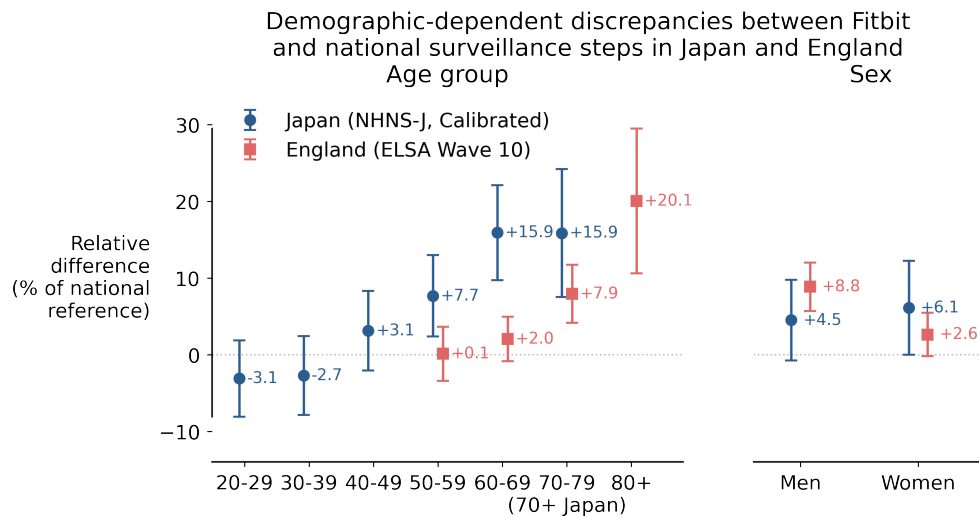


Figure 2 | Relative differences between marginal mean step counts increased with age in both countries. Estimates in Japan were adjusted by an external calibration equation from a device comparison study in participants aged 21–50, and over this age range, relative differences sit close to 0% ([Nakagata et al., 2022](#)). The English Longitudinal Study of Ageing accelerometer substudy deployed a more comparable wrist-worn device. The relative difference was 0.1% for participants aged 50–59 and was lower than the corresponding difference seen in Japan, though it also increased with age.

marginal means over 10-year age groups in Japan resulted in a MAPE of 9.1% (95% CI: 6.6–12.0). This suggests that differences between Fitbit and national estimates are smaller in older participants when device hardware and placement is similar, though small differences persist between the estimates.

The difference between the Fitbit and national surveillance estimates increased with age in both countries. In Japan, the MPD increased from -3.1% (95% CI: -8.1–1.9) in participants aged 20–29 to 15.9% (95% CI: 7.5–24.2) in participants aged 70+. In England the MPD increased from 0.1% (95% CI: -3.4–3.6) in participants aged 50–59 to 20.1% (95% CI: 10.6–29.5) in participants aged 80+. Comparing common age groups, we observed smaller differences between the Fitbit cohort and English national surveillance estimate than in Japan. In participants aged 50–59, the differences were +0.1% (95% CI: -3.4%–3.6%) in England compared with +7.7% (95% CI: 2.4–13.0) in Japan, and for participants aged 60–69, 2.0% (95% CI: -0.9–4.9) in England compared with +15.9% (95% CI: 9.7–22.1) in Japan. Overall, these results indicate that Fitbit step count estimates are within a few percentage points of national estimates in adults aged 20–50 in Japan and 50–70 in England, but are higher than national estimates in older adults in both countries (Supplementary Table 2).

Beyond comparing the magnitude of the step count estimates in Japan, we can compare their trends over time. Both Fitbit and NHNS-J age and gender-standardised cohorts showed a decline in mean daily steps over 2015–2023 of 8.6% and 6.6% respectively. NHNS-J data collection was paused during the pandemic (2020–2021). However, data from the Fitbit cohort was available throughout this period, and showed a 4.5% decline from 2019 to 2020.

### 2.3. Comparing trends in self-reported sufficient physical activity to Fitbit-measured step counts

After showing that Fitbit data were comparable to snapshot national surveillance estimates, we assessed whether continuous, multiyear Fitbit data

could reliably estimate population-level trends in step counts. Whilst there are few other countries that offer nationally-representative, device-based step estimates, there are prior estimates of national and regional trends in insufficient physical activity across 197 countries and territories over several decades based on self-reported physical activity surveys (Strain et al., 2024). Twenty-eight countries had at least 5 years of mean daily step estimates and sufficient demographic coverage, defined as at least 30 Fitbit users per combination of 10-year age group and gender (Supplementary Table 3). These 28 countries are predominantly in the Global North.

Country-level changes in Fitbit step counts were associated with changes in the prevalence of self-reported sufficient physical activity. Linear trends fitted to Fitbit data from 2015 to 2022 and extrapolated to 2010 explained 47% of the between-country variation in prevalence changes among women and 50% among men (Figure 3). Each 1% increase in mean Fitbit steps was associated with a 1.14% increase in the prevalence of self-reported sufficient physical activity among women (SE=0.239,  $p<0.001$ ) and a 0.79% increase among men (SE=0.156,  $p<0.001$ ). For instance, in Ireland there was an estimated 17.1% increase in the prevalence of women meeting physical activity guidelines from 64.8% in 2010 to 75.9% in 2022. Similarly, we saw an increase in mean daily steps in women using Fitbits in Ireland over 2015–2022 of 10.1 steps per day, year over year. On the other hand, we also observed discrepancies between these estimates. In South Africa there was an estimated 6.1% increase in the prevalence of women meeting physical activity guidelines, whereas we saw an average decrease in mean daily steps of 216.7 steps per year.

### 2.4. Assessing Fitbit data to fill gaps in national surveillance

As annual health surveys are costly and burdensome, we tested whether continuous Fitbit data could help fill gaps between national surveillance waves. We therefore simulated if using national Fitbit data might be more accurate than relying on older national surveillance estimates. To evaluate this using annual device-based surveys from

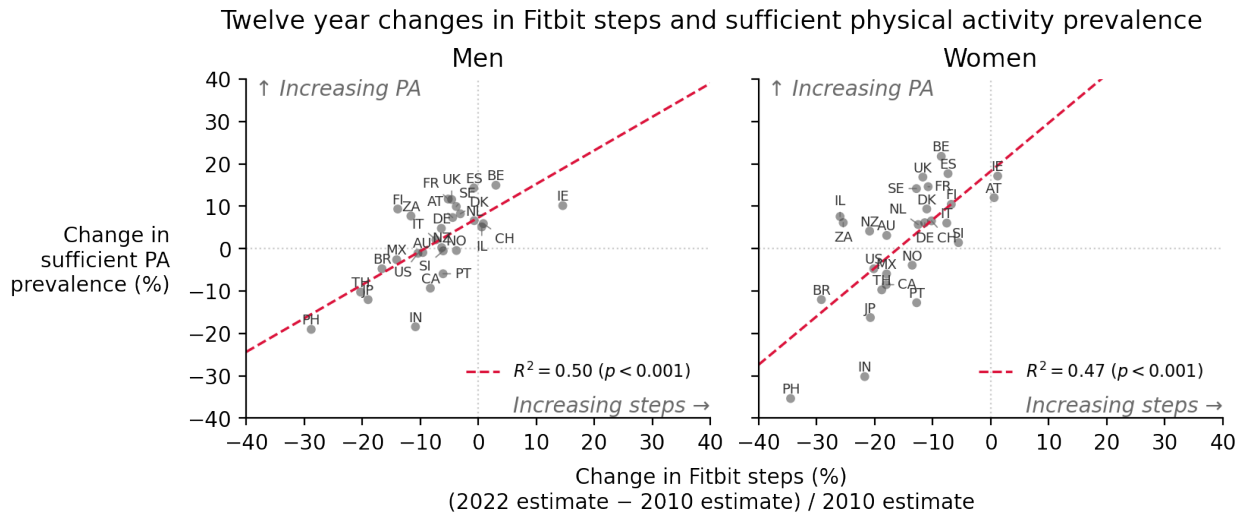


Figure 3 | Extrapolated 12-year relative changes in Fitbit mean step counts showed moderate-to-strong agreement with 12-year relative changes in the prevalence of self-reported sufficient physical activity (2010-2022; [Strain et al. \(2024\)](#)). A mapping from country codes to country names is available in Supplementary Table 3.

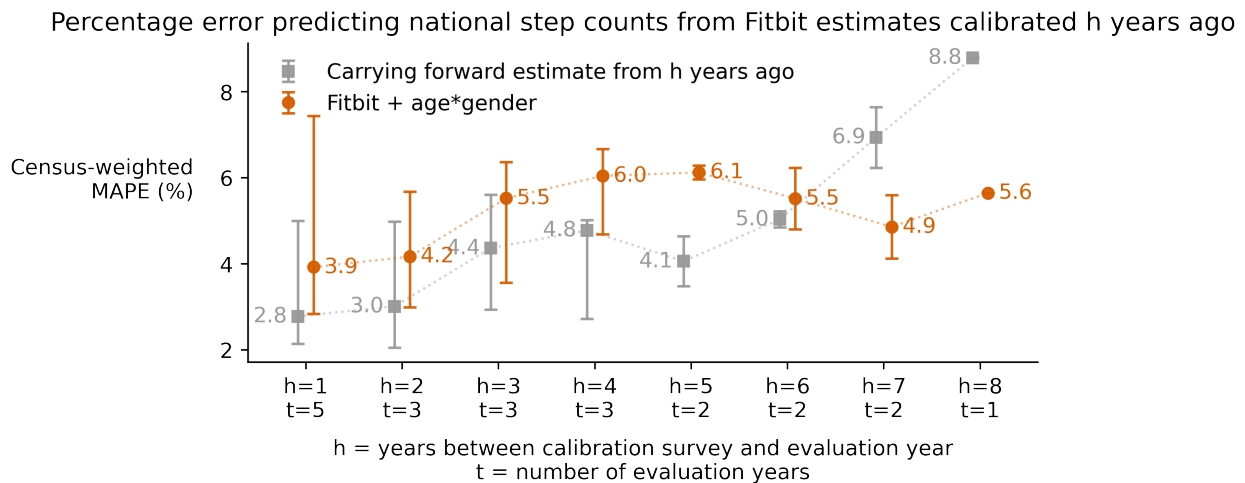


Figure 4 | MAPE in predicting national step counts in Japan from contemporary Fitbit estimates, with an OLS model, adjusting for age group and gender, trained on concurrent data from  $h$  years prior to the evaluation year. The baseline strategy of carrying forward estimates from  $h$  years prior performs better for small lags. The number of evaluation years for each lag  $h$  is indicated with  $t$ , which jumps from 5 to 3 due to missing estimates during the COVID-19 pandemic. Central markers represent the median across evaluation years, with whiskers representing the minimum and maximum observed differences (point estimates shown when  $t = 1$ ).

Japan, we simulated scenarios where national step estimates would not be collected for  $h$  years (Figure 4). Initially, a simple approach of carrying forward the previous year's estimates could be sufficiently accurate for up to 7 years. However, after 7 years the inclusion of age and sex standardised Fitbit data helped improve national step count estimates (Median MAPE at 7 years 4.9% vs. 6.9% steps/day; Supplementary Table 5).

### 3. Discussion

Fitbit mean daily step estimates showed close agreement with national surveillance estimates in younger and middle-aged adults in Japan (20–50 years: MPD =  $-0.3\%$ ) and England (50–69 years: MPD =  $0.1\text{--}2.0\%$ ), and 12-year changes in mean daily steps across 28 countries showed a moderate-to-strong association with changes in the prevalence of people meeting physical activity guidelines ( $R^2 = 0.47\text{--}0.50$ ), despite demographic imbalances. However, Fitbit estimates substantially overestimated mean daily steps in older adults in both Japan (70+ years: MPD =  $15.9\%$ ), and England (80+ years: MPD =  $20.1\%$ ). Using data from Japan, we found that the inclusion of national Fitbit data could improve national step-count estimates when a previous survey is 7 or more years old. Finally, the 28 countries with sufficient demographic coverage to derive these longitudinal estimates are predominantly in the Global North.

Dismissing consumer wearable cohorts solely on the basis of non-representative sampling would overlook valuable epidemiological signals. While Fitbit users tended to be younger, aligning with previous observations that device adoption favours younger individuals (Bietz et al., 2016; Pontin et al., 2021), we also found that the Fitbit cohort can yield relatively accurate population step estimates in younger and middle-aged adults in Japan and England, and has some predictive power for longitudinal trends across 28 countries. While more accuracy in adults aged over 70 is important, we note that current World Health Organisation priorities for national surveillance are in adults under the age of 70 (World Health Organisation, 2026).

Comparing demographic and device-related sources of discrepancy in Japan highlighted the relatively large impact of device differences. Whilst adjusting for device-related measurement differences helped reduce the differences in estimates, we still observed age and gender varying differences between the estimates, either due to differences in the cohort, or limitations with the applied calibration equation (Figure 2). Age-related differences in device estimates have previously been observed in a study comparing wrist-worn ActiGraph GT3X and AS-200-measured steps in the same population (Nakagata et al., 2024), as well as in a systematic review of wrist versus hip-worn ActiGraphs (Toth et al., 2024). For this reason, the comparison in England was informative since we were able to compare estimates in older adults wearing similar devices. Indeed, we observed lower relative differences in the marginal means in each age group in the UK compared to Japan, though there were still age-related differences.

This study used one of the largest consumer wearable dataset, and was able to compare it to national physical activity estimates from three different sources. As far as possible, we tried to match the temporal sampling and data processing procedures used in these studies. The comparison in Japan was limited by the availability of comparison studies to harmonise the AS-200 and Fitbit step counts. Step counts were aggregated across different Fitbit devices, device firmware and mobile application software, and shifts in the distribution of these over time could impact the estimates. Additionally, we were unable to make adjustments based on covariates such as socioeconomic status, location and BMI, which might further explain differences in steps between these cohorts. For the ELSA comparison, we contrasted estimates from the UK (2022 population: 67.6 million) with those from England (57.1 million, or 84.5% of the UK population Office for National Statistics (2024)), noting that ELSA does not include participants from Scotland, Wales, or Northern Ireland.

Consumer wearable data could extend national physical activity surveillance by providing estimates between survey waves, at least in settings

with high device adoption such as Japan and England. Their limited reach in the Global South, however, highlights an important equity gap. Expanding validation across countries and cohorts, while investing in representative wearable-based surveillance, will be critical to realise the potential of consumer devices for global population health monitoring.

## 4. Methods

### 4.1. Study populations and data processing

#### 4.1.1. Fitbit cohort

De-identified daily step counts and self-reported demographic attributes, including birth year, gender and regional time-zone, were obtained from Fitbit users who provided informed consent for secondary data analysis in scientific research. Ethical review for this study was exempted by the Institutional Review Board (IRB) of WCG (Ref. 20254444) as the research involved the secondary analysis of existing data (archival, de-identified records) that are not individually identifiable, consistent with federal regulation 45 CFR S46.104(d)(4).

#### 4.1.2. Japanese National Health and Nutrition Survey (NHNS-J)

Pedometer-measured step statistics in Japan were obtained from the e-stat portal from 2015 to 2023 (Ministry of Health, Labour and Welfare, 2026). Step estimates are available for 2024, though were derived using a different device and data-cleaning procedures, thus we excluded these estimates. Stratum-level summary statistics, including mean daily steps, sample standard deviations and participant counts were extracted across 10-year age strata (20-29, 30-39, ..., 60-69, 70+ years), for male, female and combined participants. Participants were asked to wear the devices during a representative day during a selected week in November for the full day and record the daily step count value themselves, since the pedometer did not have a memory function.

#### 4.1.3. English Longitudinal Study of Ageing (ELSA) Substudy

To evaluate Fitbit estimates against an independent wearable study, marginal mean step counts were obtained from the ELSA physical activity monitoring substudy conducted between June 2021 and March 2023. A representative sample of approximately 5,000 adults aged 50+ years in England wore an accelerometer on the wrist for 8 consecutive days and nights. Marginal mean steps and confidence intervals were available for 10-year age brackets (50-59, 60-69, 70-79, 80+ years), and for women and men (Brocklebank et al., 2026).

#### 4.1.4. United Nations World Population Prospects

The number of people in different demographic groups in Japan, the United Kingdom and countries with a large proportion of Fitbit users were obtained from the UN World Population Prospects estimates (United Nations, Department of Economic and Social Affairs, Population Division, 2024). Estimates were available across 5-year age groups, which we aggregated into 10-year age groups to align with the national studies.

#### 4.1.5. Device hardware and processing

Fitbit devices spanned several generations and form factors. Onboard tri-axial MEMS accelerometers were used to derive steps counts using a proprietary step-counting algorithm. The NHNS-J deployed waist-worn, spring-lever pedometers (AS-200, Yamasa Tokei Keiki Co., Ltd., Tokyo, Japan). The ELSA substudy deployed wrist-worn triaxial accelerometers (Axivity AX3, Axivity Ltd., Newcastle upon Tyne, UK) and derived counts using the open-source, validated OxWearables algorithm (Small et al., 2024).

### 4.2. Data cleaning and inclusion criteria

To mitigate against confounding introduced by variable non-wear, we excluded days where devices were worn for less than 80% of the 24 hour day based on the number of valid heart-rate readings, which acted as a proxy for the device being

on the wrist. Readings were aggregated by participant before computing stratum-level statistics to ensure that statistics are not skewed towards more active users. In line with the data collection and processing procedures in the NHNS-J, daily step count readings were only selected from days in November between 2015 and 2023, from Monday to Saturday (inclusive). Step counts  $< 100$  or  $\geq 50,000$  were excluded prior to aggregation (Kikuchi et al., 2025). In the United Kingdom, we selected participants aged 50+ during the ELSA wearable sub-study collection window (June 2021 – March 2023), and took the median daily step count per participant over the window (Brocklebank et al., 2026). Globally, countries were eligible if they had at least 5 years of step data between 2015 and 2022 with complete demographic coverage, defined as  $\geq 30$  active users in every 10-year age–gender stratum ( $N = 28$  countries). Participant daily step counts were aggregated per year by taking the mean.

### 4.3. Statistical analysis

#### 4.3.1. Demographic characteristics of Fitbit users

In order to assess representativeness of and growth in Fitbit users, we counted the number of unique users per year, 10-year demographic group and gender. Let  $N_{i,g,t}$  denote the number of active Fitbit users in age group  $i$ , gender  $g$ , and calendar year  $t$ . The proportion within each stratum is  $p_{i,g,t} = N_{i,g,t} / \sum_{i',g'} N_{i',g',t}$ . National census population weights  $w_k$  for stratum  $k = (i, g)$  were derived from UN population data and normalized such that  $\sum_k w_k = 1$ . Since ELSA data collection happened over several years, we compared the number of unique Fitbit participants to the average number of people in each demographic group from the population estimates over the ELSA data collection period.

#### 4.3.2. Differences in step counts in Japan and the UK

To quantify differences between Fitbit and national step count estimates, we used three main metrics. Let  $F_{k,t}$  denote the mean daily step count from the Fitbit cohort for demographic stratum

$k$  in year  $t$ , and let  $Y_{k,t}$  denote the corresponding national estimate over  $T$  concurrent years. The Mean Absolute Error (MAE) was computed as,

$$\text{MAE} = \frac{1}{T} \sum_{t=1}^T \sum_k w_k |F_{k,t} - Y_{k,t}|.$$

The Mean Absolute Percentage Error (MAPE), which expresses these absolute differences relative to the national estimates, was computed as,

$$\text{MAPE} = \frac{1}{T} \sum_{t=1}^T \sum_k w_k \left| \frac{F_{k,t} - Y_{k,t}}{Y_{k,t}} \right| \times 100\%.$$

The Mean Percentage Difference (MPD) omits the absolute value operation to measure average bias, and was computed as

$$\text{MPD} = \frac{1}{T} \sum_{t=1}^T \sum_k w_k \left( \frac{F_{k,t} - Y_{k,t}}{Y_{k,t}} \right) \times 100\%.$$

To evaluate statistical uncertainty around the multi-year metrics in Japan, we performed year-clustered bootstrap sampling of calendar years with replacement ( $B = 10,000$  iterations, retaining all 12 demographic strata within sampled years). The limited number of concurrent survey waves and potential temporal autocorrelation across survey years are limitations with this approach.

To address measurement differences between wrist-worn Fitbits measured steps and AS-200 pedometer measured steps, we applied the following equation:

$$F_{k,t}^{(\text{adj})} = \frac{F_{k,t} - 4152}{0.8}.$$

This was taken from an external comparison study comparing concurrent wear of the Fitbit Flex and AS-200 pedometer in 19 participants aged 21–50, conducted in free-living conditions (Nakagata et al., 2022).

Cross-sectional estimates in ELSA were only available as marginal means across age groups (4 estimates) and gender (2 estimates). To compare Fitbit estimates to ELSA estimates, we performed ordinary least squares (OLS) regression with participant median steps as the outcome and age

group and gender as categorical variables. To compare the differences seen in England to those seen in Japan, we computed marginal means over Fitbit estimates in Japan and NHNS-J estimates by fitting similar regression models with stratum-level step estimates as the outcome and age, gender, and calendar year as covariates.

The marginal mean  $\hat{\mu}_m$  can be expressed as a linear combination of the fitted regression coefficients  $\hat{\beta} \in \mathbb{R}^{p \times 1}$  as  $\hat{\mu}_m = c^\top \hat{\beta}$ , where  $c \in \mathbb{R}^{p \times 1}$  is a contrast vector obtained from averaging over all values of non-target covariates while holding the value of a target covariate fixed. By the central limit theorem, the marginal mean is asymptotically normally distributed with mean  $c^\top \beta$  and standard error  $SE_m = (c^\top \Sigma_{\hat{\beta}} c)^{1/2}$ , where  $\Sigma_{\hat{\beta}} \in \mathbb{R}^{p \times p}$  is the covariance matrix of the fitted regression parameters, thus enabling us to compute confidence intervals around marginal means.

For marginal means across age groups or gender, we computed the Mean Percentage Difference (MPD) and Mean Absolute Percentage Error (MAPE) between the Fitbit ( $\hat{\mu}_{FB,m}$ ) and national ( $\hat{\mu}_{Ref,m}$ ) marginal means. Using the Delta method, the sampling distribution of the MPD is asymptotically normally distributed with mean  $100 \left( \frac{\mu_{FB,m}}{\mu_{Ref,m}} - 1 \right)$  and standard error:

$$SE(MPD_m) \approx 100 \left( \frac{\hat{\mu}_{FB,m}}{\hat{\mu}_{Ref,m}} \right) \times \left[ \left( \frac{SE_{FB,m}}{\hat{\mu}_{FB,m}} \right)^2 + \left( \frac{SE_{Ref,m}}{\hat{\mu}_{Ref,m}} \right)^2 \right]^{1/2}$$

For the Mean Absolute Error (MAE), the linear difference  $\hat{\mu}_{FB,m} - \hat{\mu}_{Ref,m}$  is normally distributed with mean  $\mu_{FB,m} - \mu_{Ref,m}$  and exact standard error  $SE(MAE_m) = (SE_{FB,m}^2 + SE_{Ref,m}^2)^{1/2}$ . Standard errors for the MAPE equal  $SE(MPD_m)$ , with confidence intervals bounded at zero ( $\geq 0$ ).

In order to summarise differences across both Japan and England, we took the average of the differences in the marginal means across age groups. To construct confidence intervals for these composite summaries, we drew 10,000 parameter realisations from the sampling distribution of the fitted regression parameters and recorded the 2.5th and 97.5th percentiles of the resulting composite metrics.

Mean step count declines between 2015 and 2023 were calculated from annual demographic-standardised means. The relative change over this observation period was obtained by taking the differences between the 2023 and 2015 demographic standardised means, and dividing by the 2015 mean.

#### 4.3.3. Comparing trends in sufficient physical activity to step counts

Strain et al. (2024) report model-adjusted estimates of the prevalence of insufficient physical activity in different countries in 2010 and 2022. The relative change in sufficient physical activity was obtained from these estimates by looking at the difference between the 2022 and 2010 estimates relative to the 2010 estimates. This was informed by predicted trends in the prevalence of insufficient physical activity based on a multi-level Bayesian model and national estimates from physical activity questionnaires (Strain et al., 2024).

Annual world-standardised mean daily steps for each eligible country ( $N = 28$  countries meeting complete demographic coverage) and gender were modelled as continuous linear functions of calendar year over the observation window 2015–2022. To align with the 12-year physical activity surveillance window of Strain et al. (2024), estimated 12-year relative percentage changes from 2010 to 2022 were extrapolated. The cross-national relationship between wearable step count shifts and population survey trends across the 28 countries was evaluated using ordinary least squares regressions specified separately for men and women, yielding variance explained, regression slopes, standard errors, and baseline intercepts.

#### 4.3.4. Assessing Fitbit data to fill gaps in national surveillance

National device-based step estimates, if available at all have been collected for only a single or a couple of years in countries like Denmark and the Czech Republic (de Wolf et al., 2024). We evaluated whether annual Fitbit estimates can provide contemporary estimates in these settings by calibrating models against historical national

estimates.

We used stratum-level NHNS-J data from 2015 to 2023, and trained models on concurrent national and Fitbit data up to  $h$  years before an evaluation year. Whilst the model is able to use Fitbit estimates for the evaluation year, it only has access to concurrent national and consumer estimates from  $h$  years before. The lag  $h$  between concurrent seen estimates and the evaluation year was varied from 1 to 8 years, where a lag of 8 years involved training models on concurrent data from 2015 and then testing its performance on unseen NHNS-J estimates from 2023. Smaller lags enabled multiple train-test pairs. For instance, a lag of  $h = 4$  years led to the following train-test pairs  $(\text{train}, \text{test}) \in \{(\{2015\}, \{2019\}), (\{2015, 2016, 2017, 2018\}, \{2022\}), (\{2015, 2016, 2017, 2018, 2019\}, \{2023\})\}$ . Performance on evaluation years was assessed by computing the stratum-level MAE, MPD and MAPE.

## 5. Data availability

Fitbit data cannot be made publicly available due to consent provided by users which grants permission for use of this data for research purposes but not for public release. Ethical approval for the use of this data was granted by the Western Institutional Review Board-Copernicus Group IRB, which classified the research as exempt under federal regulation 45 CFR 46.104(d)(4). Upon paper acceptance, we will provide an exemplary synthetic data -set, in order to illustrate the methods used in this paper, so that they can be replicated on other dataset.

Annual, stratum-level mean daily step estimates from the NHNS-J were obtained from the e-stat portal ([Ministry of Health, Labour and Welfare, 2026](#)). These data are publicly available and free of copyright restrictions for academic reuse. Cross-sectional, marginal mean step estimates from the ELSA physical activity monitoring substudy were obtained from Table 2 of [Brocklebank et al. \(2026\)](#), which is available under a CC BY 4.0 International license. Finally, estimates of the prevalence of insufficient physical activity

were taken from the supplementary materials of [Strain et al. \(2024\)](#), which is distributed under a CC BY license.

## 6. Code availability

A reproduction of the analysis on synthetic data will be made available with the paper upon publication.

## 7. Competing interests

This study was funded by Google. DR, JP, RH, YC are Google employees and may own stock as part of the standard compensation package. AS and KV worked as student researchers at Google while this research was conducted.

## 8. Author contributions

AS conceived and designed the study, conducted the analysis and drafted the manuscript. YC, JP and KV contributed with data curation and extraction. AS, AD, DR, JP and KV advised on methodology, interpretation of the data analysis and discussion. All authors contributed to reviewing and revising the manuscript. DR and AD contributed to project supervision.

## 9. Acknowledgements

We wish to thank Shiho Amagasa and Shigeru Inoue for clarifying the NHNS-J data collection and processing procedures, and to Shiho in particular for compiling the yearly step-data from the e-Stats portal. Additionally, we extend our gratitude to Soren Brage for his comments on an early presentation of these results.

## 10. Research funding

AS and KV were funded as student researchers while at Google. AS is supported by the EPSRC Centre for Doctoral Training in Health Data Science (EP/S02428X/1). AD's research team is supported by a range of grants from the Wellcome Trust [223100/Z/21/Z, 227093/Z/23/Z],

- 656 Novo Nordisk, Swiss Re, Boehringer Ingelheim,  
657 National Institutes of Health's Oxford Cambridge  
658 Scholars Program, EPSRC Centre for Doctoral  
659 Training in Health Data Science (EP/S02428X/1),  
660 British Heart Foundation Centre of Research Ex-  
661 cellence (grant number RE/18/3/34214), and  
662 funding administered by the Danish National Re-  
663 search Foundation in support of the Pioneer Cen-  
664 tre for SMARTbiomed.
- 665 **References**
- 666 M. J. Bietz, C. S. Bloss, S. Calvert, J. G. Godino,  
667 J. Gregory, M. P. Claffey, J. Sheehan, and  
668 K. Patrick. Opportunities and challenges in the  
669 use of personal health data for health research.  
670 *J. Am. Med. Inform. Assoc.*, 23(e1):e42–8, Apr.  
671 2016.
- 672 L. Brocklebank, A. Steptoe, M. Bloomberg, and  
673 A. Doherty. Social, economic, and environ-  
674 mental disparities in device-measured 24-hour  
675 movement behaviours in a nationally represen-  
676 tative cohort of older english adults. *J. Gerontol.*  
677 *A Biol. Sci. Med. Sci.*, (glag197):glag197, Aug.  
678 2026.
- 679 F. C. Bull, S. S. Al-Ansari, S. Biddle, K. Borodulin,  
680 M. P. Buman, G. Cardon, C. Carty, J.-P. Chaput,  
681 S. Chastin, R. Chou, P. C. Dempsey, L. DiPietro,  
682 U. Ekelund, J. Firth, C. M. Friedenreich, L. Gar-  
683 cia, M. Gichu, R. Jago, P. T. Katzmarzyk, E. Lam-  
684 bert, M. Leitzmann, K. Milton, F. B. Ortega,  
685 C. Ranasinghe, E. Stamatakis, A. Tiedemann,  
686 R. P. Troiano, H. P. van der Ploeg, V. Wari, and  
687 J. F. Willumsen. World health organization  
688 2020 guidelines on physical activity and seden-  
689 tary behaviour. *Br. J. Sports Med.*, 54(24):1451–  
690 1462, Dec. 2020.
- 691 I. de Wolf, A. Elevelt, F. van Nassau, V. Toe-  
692 poel, E. de Hollander, M. E. Kompier, A. Luiten,  
693 B. Schouten, G. C. W. Wendel-Vos, and H. P.  
694 van der Ploeg. Comparing national device-  
695 based physical activity surveillance systems:  
696 a systematic review. *Int. J. Behav. Nutr. Phys.*  
697 *Act.*, 21(1):67, July 2024.
- 698 H. Kikuchi, N. Fukushima, S. Amagasa, and S. In-  
699 oue. Temporal trends and projected daily  
step count from 213,924 pedometer data in  
a nationally representative japanese popula-  
tion. *Med. Sci. Sports Exerc.*, 57(8):1763–1768,  
Aug. 2025.
- 700 J. L. Mair, L. D. Hayes, A. K. Campbell, and  
701 N. Sculthorpe. Should we use activity tracker  
702 data from smartphones and wearables to un-  
703 derstand population physical activity patterns?  
704 *J. Meas. Phys. Behav.*, 5(1):3–7, Mar. 2022.
- 705 A. E. Mason, F. M. Hecht, S. K. Davis, J. L. Na-  
706 tale, W. Hartogensis, N. Damaso, K. T. Clay-  
707 pool, S. Dilchert, S. Dasgupta, S. Purawat, V. K.  
708 Viswanath, A. Klein, A. Chowdhary, S. M. Fisher,  
709 C. Anglo, K. Y. Puldon, D. Veasna, J. G. Prather,  
710 L. S. Pandya, L. M. Fox, M. Busch, C. Gior-  
711 dano, B. K. Mercado, J. Song, R. Jaimes, B. S.  
712 Baum, B. A. Telfer, C. W. Philipson, P. P. Collins,  
713 A. A. Rao, E. J. Wang, R. H. Bandi, B. J. Choe,  
714 E. S. Epel, S. K. Epstein, J. B. Krasnoff, M. B.  
715 Lee, S.-W. Lee, G. M. Lopez, A. Mehta, L. D.  
716 Melville, T. S. Moon, L. R. Mujica-Parodi, K. M.  
717 Noel, M. A. Orosco, J. M. Rideout, J. D. Ro-  
718 bishaw, R. M. Rodriguez, K. H. Shah, J. H. Sie-  
719 gal, A. Gupta, I. Altintas, and B. L. Smarr. De-  
720 tection of COVID-19 using multimodal data  
721 from a wearable device: results from the first  
722 TemPredict study. *Sci. Rep.*, 12(1):3463, Mar.  
723 2022.
- 724 Ministry of Health, Labour and Welfare. Na-  
725 tional health and nutrition survey. [https://www.e-stat.go.jp/stat-search/  
726 files?page=1&toukei=00450171&  
727 tstat=000001041744&cycle=7&cycle\\_  
728 facet=cycle&metadata=1&data=1](https://www.e-stat.go.jp/stat-search/files?page=1&toukei=00450171&tstat=000001041744&cycle=7&cycle_facet=cycle&metadata=1&data=1), 2026.  
729 Accessed: 2026-5-7.
- 730 T. Nakagata and R. Ono. Data resource profile: Ex-  
731 ercise habits, step counts, and sedentary behav-  
732 ior from the national health and nutrition sur-  
733 vey in japan. *Data Brief*, 53(110103):110103,  
734 Apr. 2024.
- 735 T. Nakagata, H. Murakami, R. Kawakami,  
736 J. Tripette, S. Nakae, Y. Yamada, K. Ishikawa-  
737 Takata, S. Tanaka, and M. Miyachi. Step-count  
738 outcomes of 13 different activity trackers: Re-  
739 sults from laboratory and free-living experi-  
740 ments. *Gait Posture*, 98:24–33, Oct. 2022.

- 746 T. Nakagata, Y. Yamada, M. Taniguchi, H. Nanri,  
747 M. Kimura, M. Miyachi, and R. Ono. Com-  
748 parison of step-count outcomes across seven  
749 different activity trackers: a free-living experi-  
750 ment with young and older adults. *BMC Sports*  
751 *Sci. Med. Rehabil.*, 16(1):156, July 2024.
- 752 Office for National Statistics. Population  
753 estimates for the UK, england, wales, scot-  
754 land, and northern ireland - office for  
755 national statistics. [https://www.ons.gov.](https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/populationestimates/bulletins/annualmidyearpopulationestimates/mid2022)  
756 [uk/peoplepopulationandcommunity/](https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/populationestimates/bulletins/annualmidyearpopulationestimates/mid2022)  
757 [populationandmigration/](https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/populationestimates/bulletins/annualmidyearpopulationestimates/mid2022)  
758 [populationestimates/bulletins/](https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/populationestimates/bulletins/annualmidyearpopulationestimates/mid2022)  
759 [annualmidyearpopulationestimates/](https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/populationestimates/bulletins/annualmidyearpopulationestimates/mid2022)  
760 [mid2022](https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/populationestimates/bulletins/annualmidyearpopulationestimates/mid2022), Mar. 2024. Accessed: 2026-8-14.
- 761 J. D. Omura, S. A. Carlson, P. Paul, K. B. Wat-  
762 son, and J. E. Fulton. National physical activity  
763 surveillance: Users of wearable activity moni-  
764 tors as a potential data source. *Prev. Med. Rep.*,  
765 5:124–126, Mar. 2017.
- 766 F. Pontin, N. Lomax, G. Clarke, and M. A. Morris.  
767 Socio-demographic determinants of physical  
768 activity and app usage from smartphone data.  
769 *Social Science & Medicine*, 284:114235, Sept.  
770 2021.
- 771 J. F. Sallis and B. E. Saelens. Assessment of phys-  
772 ical activity by self-report: status, limitations,  
773 and future directions. *Res. Q. Exerc. Sport*, 71  
774 Suppl 2(sup2):1–14, June 2000.
- 775 R. J. Shephard. Limits to the measurement of  
776 habitual physical activity by questionnaires. *Br.*  
777 *J. Sports Med.*, 37(3):197–206; discussion 206,  
778 June 2003.
- 779 S. R. Small, S. Chan, R. Walmsley, L. VON Fritsch,  
780 A. Acquah, G. Mertes, B. G. Feakins, A. Creagh,  
781 A. Strange, C. E. Matthews, D. A. Clifton, A. J.  
782 Price, S. Khalid, D. Bennett, and A. Doherty.  
783 Self-supervised machine learning to character-  
784 ize step counts from wrist-worn accelerometers  
785 in the UK biobank. *Med. Sci. Sports Exerc.*, 56  
786 (10):1945–1953, Oct. 2024.
- 787 T. Strain, K. Wijndaele, M. Pearce, and S. Brage.  
788 Considerations for the use of consumer-grade  
789 wearables and smartphones in population  
790 surveillance of physical activity. *J. Meas. Phys.*  
791 *Behav.*, 5(1):8–14, Mar. 2022.
- 792 T. Strain, S. Flaxman, R. Guthold, E. Semenova,  
793 M. Cowan, L. M. Riley, F. C. Bull, G. A. Stevens,  
794 and Country Data Author Group. National, re-  
795 gional, and global trends in insufficient physi-  
796 cal activity among adults from 2000 to 2022:  
797 a pooled analysis of 507 population-based sur-  
798 veys with 5·7 million participants. *Lancet Glob.*  
799 *Health*, 12(8):e1232–e1243, Aug. 2024.
- 800 L. Toth, A. E. Paluch, D. R. Bassett, Jr, E. Rees-  
801 Punia, E. M. Eberl, S. Park, and K. R. Evenson.  
802 Comparative analysis of ActiGraph step count-  
803 ing methods in adults: A systematic literature  
804 review and meta-analysis. *Med. Sci. Sports Ex-*  
805 *erc.*, 56(1):53–62, Jan. 2024.
- 806 United Nations, Department of Economic and So-  
807 cial Affairs, Population Division. World popu-  
808 lation prospects 2024, online edition, 2024.
- 809 World Health Organisation. Integrating wearable  
810 technology into population health monitoring  
811 systems for physical activity measurement:  
812 meeting report, utrecht, kingdom of the nether-  
813 lands, 25-27 november 2025. [https://www.](https://www.who.int/publications/i/item/B09745)  
814 [who.int/publications/i/item/B09745](https://www.who.int/publications/i/item/B09745),  
815 May 2026. Accessed: 2026-8-18.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [2026SuppSchonfeldtConsumerCohortSteps.pdf](#)