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

**Abram Schönfeldt[1,2], Katarina Vukosavljević[1,3], Justin Phillips[1], YongSuk Cho[1], Robert Harle[1,4], Shwetak Patel[1,5], Aiden Doherty[2] and Daniel Roggen[1,6]**

[1]Google Research, [2]Nuffield Department of Population Health, University of Oxford, [3]Institute of Biomedical Engineering, University of Oxford, [4]Department of Computer Science and Technology, University of Cambridge, [5]Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, USA, [6]School of Engineering & Informatics, University of Sussex

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

Stratum-level differences between NHNS-J and Fitbit mean daily steps, averaged over the 7 years of concurrent estimates, are presented in Table 1.

Differences between national and Fitbit marginal means in England and Japan are shown in Table 2. Marginal means were obtained by fitting OLS regression models with age and gender as categorical variables in England, and age, gender and calendar year as categorical variables in Japan. Predictions were then made for all combinations of categories of the covariates, and these predictions were summed over the non-target dimension.

## 2. Differences in trends in physical activity in the Global North

Countries with at least 5 years of complete sets of estimates across age and gender are tabulated in Table 3. In order to label them concisely in the main figure, we referred to them by their ISO2 codes.

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

Table 4 shows the equations of the models evaluated for filling in hypothetical gaps in national surveillance in Japan.

Table 5 and Figure 1 show the performance of the models at each lag.

## References

T. Nakagata, H. Murakami, R. Kawakami, J. Tripette, S. Nakae, Y. Yamada, K. Ishikawa-Takata, S. Tanaka, and M. Miyachi. Step-count outcomes of 13 different activity trackers: Results from laboratory and free-living experiments. *Gait Posture*, 98:24–33, Oct. 2022.

*Corresponding author(s): danielroggen@google.com*

Table 1 | Census-weighted stratum-level step count discrepancies and relative bias between Fitbit and NHNS-J national pedometer benchmarks in Japan (2015–2023). Adjusted estimates were calibrated using the free-living calibration equation from Nakagata et al. (2022).

| Age Group | Adjusted | MAE (steps/day, 95% CI) | MPD (%, 95% CI) | MAPE (%, 95% CI) |
|---|---|---|---|---|
| 20–70+ | No | 3,185.7 (2,968.6–3,390.8) | 52.8 (49.7–56.0) | 52.8 (49.7–56.0) |
| 20–70+ | Yes | 690.4 (563.6–859.0) | 7.2 (3.0–11.8) | 11.4 (9.0–14.7) |
| 20–50 | Yes | 489.1 (390.2–592.5) | -0.3 (-5.4–3.4) | 6.7 (5.3–8.1) |

Table 2 | Marginal step count discrepancies and relative bias between Fitbit and national benchmarks in Japan (NHNS-J 2015–2023) and England (ELSA Wave 10, 2021–2023). Japan estimates are calibrated using Nakagata et al. (2022); England estimates compare commercial wrist-worn Fitbit devices against research wrist-worn accelerometers without adjustment.

| Country | Subgroup | MAE (steps/day, 95% CI) | MPD (%, 95% CI) | MAPE (%, 95% CI) |
|---|---|---|---|---|
| **Gender marginals** | | | | |
| Japan | Men | 324 (0–699) | 4.5 (-0.8–9.8) | 4.5 (0.0–9.8) |
| Japan | Women | 380 (5–755) | 6.1 (-0.0–12.2) | 6.1 (0.0–12.2) |
| *Japan* | *Average* | *353 (120–614)* | *5.3 (1.4–9.4)* | *5.3 (1.8–9.4)* |
| England | Men | 722 (485–959) | 8.8 (5.7–12.0) | 8.8 (5.7–12.0) |
| England | Women | 202 (0–415) | 2.6 (-0.2–5.5) | 2.6 (0.0–5.5) |
| *England* | *Average* | *450 (298–610)* | *5.6 (3.5–7.8)* | *5.6 (3.6–7.8)* |
| **Age marginals** | | | | |
| Japan | 20-29 | 231 (0–606) | -3.1 (-8.1–1.9) | 3.1 (0.0–8.1) |
| Japan | 30-39 | 197 (0–572) | -2.7 (-7.9–2.4) | 2.7 (0.0–7.9) |
| Japan | 40-49 | 228 (0–602) | 3.1 (-2.1–8.3) | 3.1 (0.0–8.3) |
| Japan | 50-59 | 556 (181–931) | 7.7 (2.4–13.0) | 7.7 (2.4–13.0) |
| Japan | 60-69 | 1,011 (636–1,386) | 15.9 (9.7–22.1) | 15.9 (9.7–22.1) |
| Japan | 70+ | 747 (373–1,122) | 15.9 (7.5–24.2) | 15.9 (7.5–24.2) |
| *Japan* | *Average* | *534 (395–697)* | *7.6 (4.8–10.5)* | *9.1 (6.6–12.0)* |
| England | 50-59 | 12 (0–358) | 0.1 (-3.4–3.6) | 0.1 (0.0–3.6) |
| England | 60-69 | 188 (0–449) | 2.0 (-0.9–4.9) | 2.0 (0.0–4.9) |
| England | 70-79 | 588 (327–848) | 7.9 (4.1–11.7) | 7.9 (4.1–11.7) |
| England | 80+ | 1,061 (643–1,480) | 20.1 (10.6–29.5) | 20.1 (10.6–29.5) |
| *England* | *Average* | *376 (289–551)* | *5.8 (3.7–8.0)* | *5.8 (4.3–8.3)* |

Table 3 | Included countries, ISO-2 abbreviations, and annual Fitbit sample sizes in 2015 and 2022. Total number of eligible Fitbit participants meeting quality and wear-time inclusion criteria in 2015 and 2022 across the 28 countries analysed.

| ISO | Country / Territory | 2015 | 2022 |
|---|---|---|---|
| AU | Australia | 9,398 | 179,872 |
| AT | Austria | 248 | 14,157 |
| BE | Belgium | 545 | 39,822 |
| BR | Brazil | 307 | 7,669 |
| CA | Canada | 18,076 | 401,822 |
| DK | Denmark | 395 | 10,304 |
| FI | Finland | 267 | 6,218 |
| FR | France | 2,360 | 93,126 |
| DE | Germany | 4,163 | 250,176 |
| IN | India | 1,176 | 107,441 |
| IE | Ireland | 1,403 | 113,711 |
| IL | Israel | 99 | 3,397 |
| IT | Italy | 1,531 | 53,170 |
| JP | Japan | 1,958 | 197,182 |
| MX | Mexico | 946 | 27,618 |
| NL | Netherlands | 973 | 75,423 |
| NZ | New Zealand | 1,664 | 35,035 |
| NO | Norway | 1,714 | 42,755 |
| PH | Philippines | 245 | 14,752 |
| PT | Portugal | 432 | 11,481 |
| SI | Singapore | 748 | 24,350 |
| ZA | South Africa | 372 | 18,864 |
| ES | Spain | 1,712 | 67,284 |
| SE | Sweden | 1,188 | 45,900 |
| CH | Switzerland | 1,186 | 34,807 |
| TH | Thailand | 373 | 11,718 |
| UK | United Kingdom | 25,907 | 707,676 |
| US | United States of America | 183,089 | 2,155,040 |

Table 4 | Equations for each of the models of mean daily steps for strata in the NNHNS-J cohort, $Y$, as a function of mean daily steps in the Fitbit cohort, $F$, age, $A$, gender, $G$.

| Model | Equation |
|---|---|
| Nakagata et al., 2022 | $Y = (F - 4152)/0.8$ |
| Fitbit + demographics | $Y = \beta_0 + \beta_1 A + \beta_2 G + \beta_3 A \cdot G + \beta_4 F$ |
| Exponential discount | $Y = F\left[G\left(\beta_0 + \beta_1 e^{\beta_2 A}\right) + (1 - G)\left(\beta_3 + \beta_4 e^{\beta_5 A}\right)\right]$ |

Table 5 | MAPE estimating national step counts in Japan from Fitbit estimates, calibrated on concurrent data from $h$ years before. Values represent the median across evaluation years, with empirical minimum and maximum bounds shown in brackets (single point estimates shown when $t = 1$).

| $h$ | $t$ | **Carry forward** | **Nakagata et al., 2022** | **Fitbit + demo** | **Exponential discount** |
|---|---|---|---|---|---|
| 1 | 5 | 2.8 (2.1–5.0) | 10.0 (7.2–19.7) | 3.9 (2.8–7.4) | 2.8 (1.3–7.5) |
| 2 | 3 | 3.0 (2.1–5.0) | 10.0 (8.9–12.5) | 4.2 (3.0–5.7) | 2.9 (1.3–5.0) |
| 3 | 3 | 4.4 (2.9–5.6) | 8.9 (8.4–10.0) | 5.5 (3.6–6.4) | 4.6 (1.8–4.9) |
| 4 | 3 | 4.8 (2.7–5.0) | 8.4 (7.2–10.0) | 6.0 (4.7–6.7) | 4.8 (3.9–5.0) |
| 5 | 2 | 4.1 (3.5–4.6) | 7.8 (7.2–8.4) | 6.1 (6.0–6.3) | 5.0 (4.8–5.2) |
| 6 | 2 | 5.0 (4.8–5.2) | 7.8 (7.2–8.4) | 5.5 (4.8–6.2) | 4.9 (4.8–5.1) |
| 7 | 2 | 6.9 (6.2–7.6) | 7.8 (7.2–8.4) | 4.9 (4.1–5.6) | 4.1 (3.4–4.7) |
| 8 | 1 | 8.8 | 7.2 | 5.6 | 5.4 |

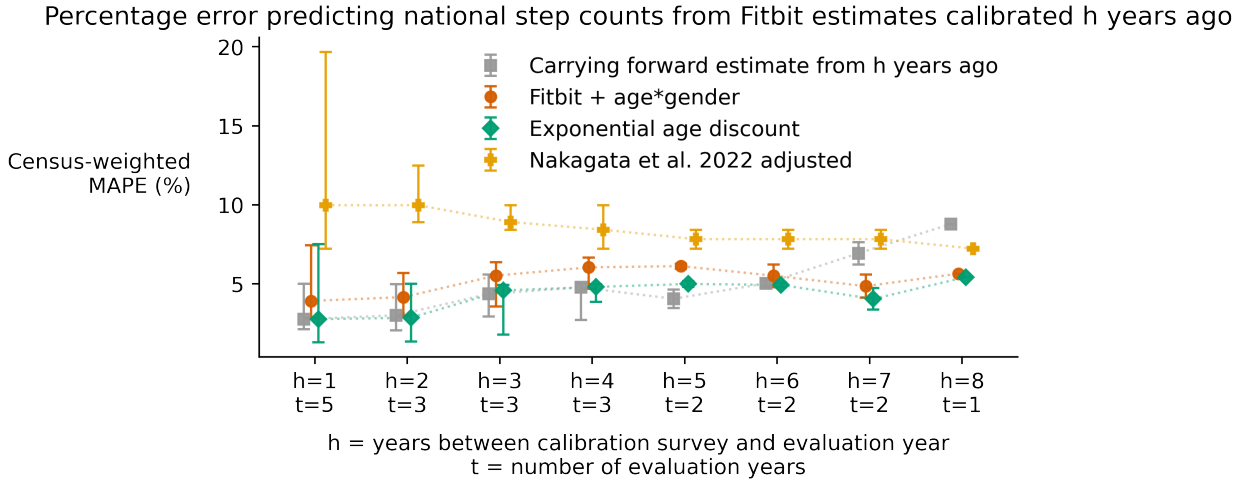


Figure 1 | Performance of model-predicted national step counts in Japan based on Fitbit estimates, with models trained on concurrent data from $h$ years prior to the evaluation year. The number of evaluation years for each lag $h$ is indicated with $t$, which jumps from 5 to 3 due to missing estimates 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 $t = 1$).