Journal for the Measurement of Physical Behaviour, (Ahead of Print)
https://doi.org/10.1123/jmpb.2021-0046
First Published Online: Feb. 2, 2022

Human Kinetics
COMMENTARY

Considerations for the Use of Consumer-Grade Wearables and Smartphones in Population Surveillance of Physical Activity

Tessa Strain, Katrien Wijndaele, Matthew Pearce, and Søren Brage
MRC Epidemiology Unit, University of Cambridge, Cambridge, United Kingdom

As smartphone and wearable device ownership increase, interest in their utility to monitor physical activity has risen concurrently. Numerous examples of the application of wearables in clinical and epidemiological research settings already exist. However, whether these devices are all suitable for physical activity surveillance is open for debate. In this article, we discuss four key issues specifically relevant to surveillance that we believe need to be tackled before consumer wearables can be considered for this measurement purpose: representative sampling, representative wear time, validity and reliability, and compatibility between devices. A recurring theme is how to deal with systematic biases by demographic groups. We suggest some potential solutions to the issues of concern such as providing individuals with standardized devices, considering summary metrics of physical activity less prone to wear time biases, and the development of a framework to harmonize estimates between device types and their inbuilt algorithms. We encourage collaborative efforts from researchers and consumer wearable manufacturers in this area. In the meantime, we caution against the use of consumer wearable device data for inference of population-level activity without the consideration of these issues.

***Keywords**:* wearable technology, activity trackers, smartwatch, device

As smartphone and wearable device ownership increases in the population (Statista, 2021), interest in their utility to monitor physical activity has risen concurrently (Wright et al., 2017). It is attractive to consider the research potential in extending the concept of the "quantified self," where individuals measure their own biological, physical, behavioral, and environmental information, to the collaborative sharing of such data (Swan, 2013). Reduced cost and participant burden, continuous monitoring over long time periods, and access to retrospective data are among the potential advantages of such methods over research-grade devices or questionnaires.

Numerous examples of the application of wearable devices in clinical and epidemiological research settings already exist (Bassett et al., 2019; Hicks et al., 2019; Strath & Rowley, 2018; Wright et al., 2017). Consumer-grade wearable devices and smartphone applications also appear to be effective components in interventions to increase activity levels (Laranjo et al., 2021). However, whether they are currently suitable for population surveillance is open for debate (Fulton et al., 2016; Mair et al., 2021; Omura et al., 2021; Strain et al., 2019). In this commentary, we will discuss four key issues that need to be tackled before consumer wearables can be considered as a potential surveillance measurement method. These are summarized in Figure 1. Wider issues more pertinent to other study designs are discussed in more detail elsewhere (Bietz et al., 2016; Cho et al., 2021; Wright et al., 2017).

Definitions

We use the term "consumer wearable devices" to cover any device that can be worn on the person that provides feedback to the user on their physical activity. This covers both monitors (usually wrist-worn smartwatches) and smartphones, and any hybrid system. Consumer wearable devices almost ubiquitously use accelerometer sensors to detect movement, but gyroscopes, magnetometers, barometers (altimeters), global positioning systems, and optical sensors (photoplethysmography) can be included to augment the inference of physical activity levels or provide additional metrics. Data may be stored on the device initially, with most offering Bluetooth connectivity to phone or tablet apps. Some also have the ability to extract and share, at a minimum, summary statistics relating to activity levels.

We define physical activity surveillance as the monitoring of physical activity behaviors in a sample representative of the target population with the aim of identifying differential levels by sociodemographic subgroups and monitoring population trends (Omura et al., 2021).

Representative Sampling

There have been extensive discussions about the merits of representative sampling for different study designs employed in epidemiological research (Richiardi et al., 2013; Rothman et al., 2013). However, all agree that when trying to describe a population, as surveillance systems aim to do, a representative sample is essential.

As one of the purported benefits of surveillance using consumer wearables is that individuals are already collecting the data (Mair et al., 2021), we will first discuss issues relating to

© 2022 The Authors. Published by Human Kinetics, Inc. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International License, CC BY 4.0, which permits unrestricted noncommercial and commercial use, distribution, and reproduction in any medium, provided the original work is properly cited, the new use includes a link to the license, and any changes are indicated. See http://creativecommons.org/licenses/by/4.0. This license does not cover any third-party material that may appear with permission in the article.

Strain (tessa.strain@mrc-epid.cam.ac.uk) is corresponding author, https://orcid.org/0000-0002-7086-1047.

Brought to you by CAMBRIDGE UNIVERSITY | Unauthenticated | Downloaded 02/04/22 10:31 AM UTC

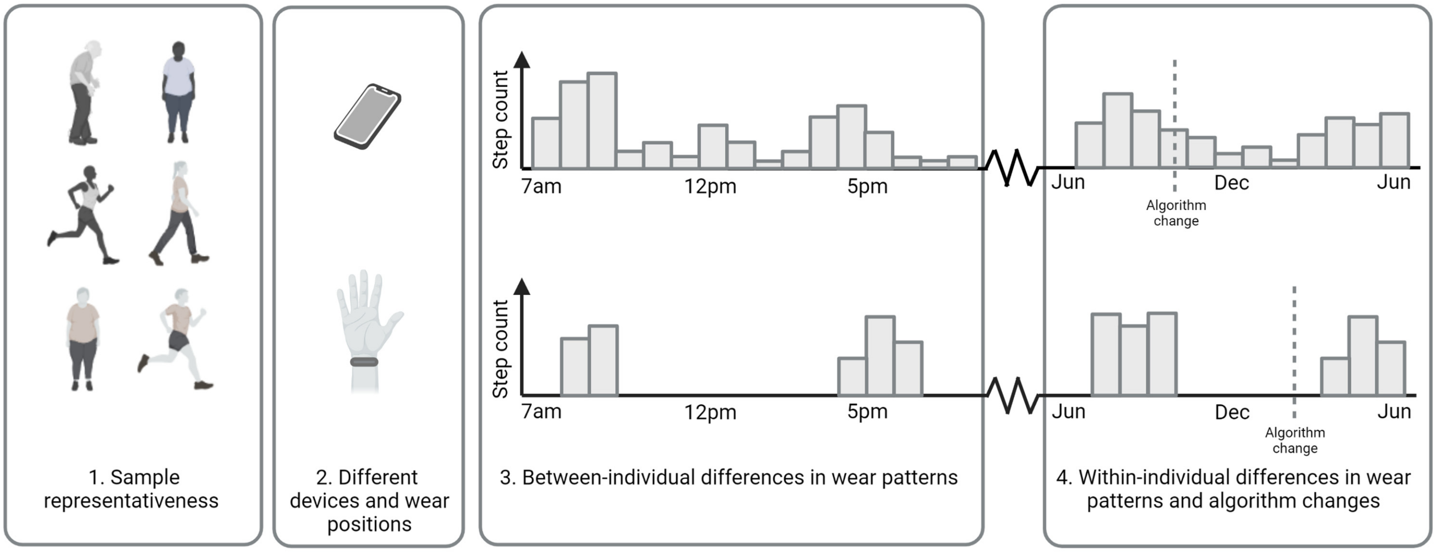

Figure 1 — An overview of the four issues of using wearable devices for physical activity surveillance covered in this article. Created with biorender.com.

representativeness that may arise if a random probability sample was drawn from existing consumer wearable device owners.

Samples that do not reflect the characteristics of the wider population may yield biased prevalence estimates and associations that are not generalizable beyond that sample. Traditionally, statistical weighting procedures have been used to mitigate such biases. Weighting usually involves multiplying the observations by a factor, derived from the probability of participation, to compensate for under- or over-representation in the sample (Gray et al., 2013). The weighted sample should match the target population on the key selected characteristics, such as age or affluence. In our example of a sample consisting of existing consumer wearable device owners, we would expect younger and more affluent individuals to be over-represented, while older and less affluent individuals would likely be under-represented as these characteristics have been associated with device ownership and use (Bietz et al., 2016; Pontin et al., 2021; Strain et al., 2019). As age and affluence are also associated with activity levels, we would expect, in an unweighted sample of consumer device owners that is likely skewed toward younger age and greater affluence, physical activity prevalence estimates to be higher than the true level in the general population. Associations between activity levels and other factors would not necessarily be generalizable to the whole population, and in particular, the under-represented groups. These biases are mitigated (although unlikely ever eliminated) in weighted analyses, as the contribution of the data from under- and over-represented groups in the sample to the population-wide estimates are adjusted to reflect the sociodemographic composition of the target population.

A key assumption of weighting is that those who do participate in the research are similar to those who do not, with respect to the outcomes of interest (Gray et al., 2013). It is therefore a concern if there are very few individuals with a certain age or affluence profile in the sample, as it is more likely that they would not be representative of the wider population with those characteristics. However, even if these subgroups are well represented, the assumption may still be violated when we investigate physical activity levels. In our example, we are assuming that those in the sample and in the wider target population with similar relevant characteristics such as age and affluence, also have similar physical activity levels. Unfortunately, we know this not to be true: we know consumer device owners are more active than nonowners, even accounting for their younger age and greater affluence (Ernsting et al., 2017; Omura et al., 2017; Strain et al., 2019; Toepoel et al., 2021; Xie et al., 2020). For example, Omura et al. (2017) found the prevalence of sufficient activity to be 50% among U.S. consumer panel survey respondents, but 70% when only considering those who were current wearable device users. This is not a surprise, as a major purpose of such devices is to change and monitor activity behavior (Mercer et al., 2016). The implication is that, even if the weighted sample matches the general population with regards to the distribution of age, affluence, and other sociodemographic factors, the prevalence estimates for physical activity are likely to be higher than the true population level and associations may be biased.

This is why surveillance based on existing consumer wearable device owners is more problematic than existing surveillance using random population sampling with more traditional methods for measuring physical activity. That being said, those who agree to participate in existing surveillance systems may also give rise to biased samples, as this design is also prone to differential response rates across demographic group. However, it is unlikely that the probability of response correlates with physical activity levels to the same degree as consumer wearable device ownership, which aim to change or maintain activity behavior. In other words, responders and nonresponders to surveys with a population sampling frame are likely to be more similar in their activity levels than consumer wearable device owners and nonowners. Therefore, it is more plausible that the key assumption of weighting holds in existing surveillance systems when it comes to estimating physical activity levels for a target population.

A potential solution may be to randomly sample from the whole population, and then provide participants with a consumer wearable device without the need to return it, utilizing an online data transfer protocol. In the near future, this solution may even be more cost-efficient compared with the delivery and return of research-grade devices. It does, however, add an element of

Brought to you by CAMBRIDGE UNIVERSITY | Unauthenticated | Downloaded 02/04/22 10:31 AM UTC

both research cost and participant burden back into the equation as these would be data that most likely would not have been collected otherwise. The approach would also miss out on retrospective data collection prior to study initiation. It could also interfere with the personal monitoring for those that did own a different device if there are differences in the capabilities, output, or mechanism for data transfer. Potentially most importantly, simply owning a device does not necessarily mean individuals are motivated to wear or charge it regularly, or download and share data with researchers (Hendker et al., 2020; Patel et al., 2015).

Indeed, further biases may be introduced in the final sample of data when considering the fact that only those willing to share their device data will be represented in the data set, rather than all device owners (Turner et al., 2021). Research suggests that privacy concerns and organizational trust are key issues (Bietz et al., 2016; Hyde et al., 2020; Toepoel et al., 2021). If consumer wearable devices are to be considered as a physical activity surveillance method, then population-specific preliminary work should be undertaken to understand the above concerns and solutions explored to alleviate them.

Representative Wear Time

Let us assume that we can obtain a representative sample of participants who agree to wear a consumer wearable and share their data for surveillance purposes. As would be the case for data collected by research-grade devices, we then need to consider issues of within- and between-person differences in wear time. A key benefit of consumer wearable device data is the low participant burden, so we will assume that we would not define a specific wear protocol for study participants but merely encourage continuous wear as is now common for studies involving wearables in general and indeed recommended by manufacturers of consumer devices. We then need to consider how best to summarize an individual's activity data to reflect habitual levels.

Day of the week is known to influence activity levels (Doherty et al., 2017) and therefore traditional research methods tend to sample time in a way that would capture this variation. For example, activity questionnaires with reference time frames of past week or month, or asking study participants to wear research-grade devices for 7 days before returning the device to the research team. A potential advantage of consumer wearable devices is that measurement periods could be vastly extended, something that has been shown to improve estimates of habitual activity (Aadland & Ylvisåker, 2015; Bergman, 2018; Wareham et al., 2000). Seasonality is another determinant of activity levels (Brage et al., 2020; Strain et al., 2022) which surveillance studies tend to control for either by sampling throughout the year, or at the same period for each consecutive survey. Timing of measures would need to be considered when summarizing individuals' consumer wearable device data. After settling on the duration and longer-term timing of a measurement period, one then needs to assess whether the data recorded during that period is representative of habitual activity, because users may not wear devices continuously. Hendker et al. (2020) observed a tendency for users to wear their device during their most active periods of the day. This suggests we should be cautious in assuming that nonwear time segments are missing at random. We should hence also be careful with computing simple averages of wear-time data with adjustment for wear time or imputation of physical activity during nonwear time, which are based on this assumption. Summation techniques that attempt to limit diurnal bias by imbalances in nonwear through equal weighting of the different times of the day may reduce the magnitude of the bias, particularly the approaches based on establishing within-person patterns for which the computational basis is much larger for consumer wearables. However, wear data may still be biased toward exercise event data, considering that many consumer devices are marketed as "fitness trackers." Between-individual differences are also a concern here. Indeed, Rising et al. (2020) identified distinct usage patterns among consumer wearable device users, including "super trackers" and "non-trackers," each with their own distinct sociodemographic profile. However, these should not be insurmountable issues. Imputation methods may be developed with a greater understanding of when and why users wear their devices. Or, alternatives such as MX metrics (the X most active minutes of the day; Rowlands et al., 2019), could be investigated regarding their robustness to different wear patterns.

A purported key advantage of wearable devices is the long-term nature of data collection, potentially even retrospectively before a known study baseline. Of course, a surveillance system aiming to understand such trends would not necessarily need continuous monitoring, but even repeated measurements from the same individual are not typical in surveillance study designs. That said, such long-term trajectory data would be very useful for surveillance, meaning within-person wear time variation over a month-to-year time-frame requires attention.

A 2014 report claimed over half of U.S. consumers that previously owned an activity tracker have ceased to use it, and a third stopped using it within the first 6 months (Endeavour Partners, 2014). Meyer et al. (2017) analyzed consumer wearable device data from 104 participants of four separate physical activity research studies over 2.5 years and identified 12 use patterns including the descriptively named "try-and-drop," "slow-start," "experimenter," and "power user." This points to the different purposes for which people use wearable devices, in many cases to aid behavior change (Mercer et al., 2016) and is a reminder that the activity that is being captured may not reflect habitual behavior if there are gaps in the wear time record. Again, potential biases may be mitigated through the development of better imputation and analysis techniques but for the time being, nonrepresentative time sampling remains an important issue for data from wearables. This may be particularly relevant to research interested in activity variation due to the COVID-19 pandemic. Wearable device or smartphone wear patterns may have changed as many populations had their routines disrupted as they were subject to stay-at-home orders.

Validity and Reliability

It is standard practice in research studies to test and document the reliability and validity of all employed methods, so that the data arising from these can be interpreted accordingly. Fortunately, there are numerous studies investigating the validity and reliability of the metrics derived from at least some consumer wearables against research-grade devices in a mixture of laboratory and free-living settings (reviews include Evenson et al., 2015; Fuller et al., 2020; Feehan et al., 2018). A major challenge for the field, however, is keeping up with the new market developments, both in terms of hardware but also firmware and software upgrades (Evenson et al., 2015; Henriksen et al., 2018). As the issue of validity is not specific to surveillance, we point readers to summaries in other articles for greater discussion (Strath & Rowley, 2018; Welk et al., 2019; Wright et al., 2017). In short, the broad

Brought to you by CAMBRIDGE UNIVERSITY | Unauthenticated | Downloaded 02/04/22 10:31 AM UTC

conclusions relating to the accelerometer-derived variables are remarkably similar across devices. Step count is the most consistently performing metric: correlations with direct observations and research-grade accelerometry measures are over 0.80 (Evenson et al., 2015). However, time spent in intensity bands or energy expenditure is generally poorly quantified (Evenson et al., 2015; Feehan et al., 2018; Fuller et al., 2020). Measures of reliability across all devices and metrics are generally reported to be within acceptable levels for scientific research (Evenson et al., 2015; Feehan et al., 2018; Fuller et al., 2020). It is hard to assess whether the validity is improving with technological developments because there is also wide variety in validity study protocols (Cosoli et al., 2020). The general use of proprietary algorithms and lack of transparency in updates make this even harder (Feehan et al., 2018). Any changes in the algorithms converting raw sensor measurements to the metrics presented on the device, which are often proprietary, pose issues for validity and also within-person comparisons over time. Although we are unaware of any analysis quantifying the implications of a consumer wearable firmware update, examples from research-grade devices indicate that this is an issue worth highlighting. The low-frequency extension option in the ActiGraph GT3X (ActiGraph, Pensacola, FL), extending the range of detection at the lower end of the movement intensity spectrum, increased the overall activity volume estimates by close to 10% (Ried-Larson, 2012). Evenson et al. (2015) called for companies to make firmware updates transparent to the public, while Wright et al. (2017) offer an alternative solution as they describe how Withings have agreed to notify of updates and provide conversions.

We might speculate that the higher error in the more complex metrics, such as energy expenditure, is potentially due to the tweaking of algorithm coefficients or the addition of a default or user-provided value of body weight to the algorithm that may not be estimated with similar precision compared with that measured in a research study. Should this be the case, this provides a potential avenue for improving the accuracy of derived metrics. It also is a reminder of the importance of accurate demographic data, an essential component to physical activity surveillance for the identification of at risk subgroups. Those considering a system to collect physical activity data from consumer wearable devices should also consider how such additional information will be gathered, and how frequently it should be updated. Accessing raw sensor measurements such as raw acceleration may be another solution, bypassing algorithms designed to produce derived estimates. This measurement in particular can be calibrated to local gravity and provide brand-agnostic metrics of physical behaviors (van Hees et al., 2014). However, it is not clear that all devices and/or their cloud infrastructure store such raw information as it requires a large amount of data storage capacity, let alone challenges of data transfer from device to smartphone.

Another option may be better integration of sensor data to infer physical activity. Heart rate detection through photoplethysmography is an increasingly common feature of smartwatches, while skin temperature and galvanic response may also feasibly be detected. The measurement of these biological signals, without much on-board data processing, can also be directly validated against research- or medical-grade measurements, for which industry standards exist. Energy expenditure is more precisely predicted from combined heart rate and accelerometer sensing than accelerometry alone when using research-grade devices (Brage et al., 2015), and one would expect the same for consumer wearables, provided similar methods for individually calibrating the heart rate to energy expenditure relationship across a reasonable intensity range were employed (Brage et al., 2007). However, more free-living studies comparing consumer wearables against research-grade devices are needed for all imbedded sensor elements that provide measurements in a device, as well as their derived estimates of activity.

Differential validity between devices may be overcome with the earlier suggestion of providing the same type of devices to potential participants. However, as described above, this may create other biases and negates some of the main purported benefits of wearable device data.

A further consideration is differential validity by wear position (Düking et al., 2018). We know from studies using research-grade devices that the validity of hip- or thigh-worn accelerometers is different to those at the wrist and even between dominant and nondominant wrists (White et al., 2016, 2019), so it is likely that carrying one's phone or wearable device in a trouser or breast pocket, or even a bag, will affect measurement error. Mobile phone data are likely to be particularly challenging in this regard; studies have shown that preferred wear position for a mobile phone varies by age, sex, and culture (Cui et al., 2007; Redmayne, 2017). Such systematic biases by demographic groups are potentially problematic for surveillance.

Compatibility Between Consumer Devices

Perhaps the greatest challenge for surveillance is the incompatibility of activity estimates between consumer wearables if one were to accept any data generated from any of the thousands of device models in use today. Concurrent comparisons of multiple consumer wearable devices indicate a wide range of estimates for moderate to vigorous physical activity, total daily energy expenditure, and, to a lesser extent, step counts (Ferguson et al., 2015). However, limiting surveillance efforts to one manufacturer or model will only exacerbate the selection biases. Providing individuals with a device may overcome the issues raised in this paragraph, but we have already raised the disadvantages of this approach above.

We have previously proposed a harmonization approach to deal with research-grade device comparisons (Pearce et al., 2020). This involves using mapping equations from concurrent validity studies to understand the relationship between two measurement regimes and convert these such that they can be considered on the same scale.

Extending this harmonization approach requiring direct validation to consumer wearable device data is a daunting prospect given the sheer number of permutations of manufacturers, models, firmware/software versions, and anatomical wear positions, which may all influence the activity estimates in the stored record (Cho et al., 2021). As a result, the relevant mapping equations are often unavailable. It is also worth noting that the majority of validity studies currently undertaken on consumer wearable devices involve middle-aged adults of normal body mass index (Evenson et al., 2015): A more representative basis would be required to apply these methods in the context of population surveillance.

One solution could be a network harmonization approach which estimates relationships between devices indirectly when a concurrent direct validity study is not available (Pearce et al., 2020). All devices have some relationship with each other, and anchoring some of the most popular wearable devices (and algorithm versions) to research-grade device methods would inform the entire network allowing harmonization. On the consumer wearables side, it is likely that some of the device comparison data have

Brought to you by CAMBRIDGE UNIVERSITY | Unauthenticated | Downloaded 02/04/22 10:31 AM UTC

already been recorded and are held by companies; even those comparisons would be strengthened by publishing the results and combining them with other methodological comparisons.

It is clear that working with consumer wearable device manufacturers in this way would help realize the public health potential of the vast amounts of consumer wearable device data collected, including its use for population surveillance. There are many examples of successful large-scale collaborations with companies such as Apple and Fitbit for epidemiological research (Banks, 2020) but the true potential lies in the utilization of data stemming from multiple types following successful data harmonization. Using data directly from companies would be a new model for population surveillance, potentially avoiding contact with the individuals themselves. This poses additional challenges in obtaining additional demographic data that, as previously described, is essential. In addition, it is important to remember that while companies have an interest in advancing research and contributing to the common good, their business model is paramount (Bietz et al., 2016) and so public health researchers need to be realistic in expectations and requests. As researchers, we also need to be aware of the potential risks companies may take through data sharing (Bietz et al., 2016; Hicks et al., 2019).

Conclusion

In this commentary, we have identified four key issues as the biggest current challenges to the use of data from consumer wearable devices in the surveillance of population physical activity levels. We have suggested ways in which these challenges may be mitigated or overcome in the future and we encourage collaborative efforts from researchers and consumer wearable manufacturers in this area. In the meantime, we caution against the use of consumer wearable device data for inference of population-level activity without the consideration of these issues.

Acknowledgment

TS, KW, MP, and SB are supported by the UK Medical Research Council [grant number MC_UU_00006/4] and MP and SB are supported by the NIHR Biomedical Research Centre in Cambridge (IS-BRC-1215-20014).

References

Aadland, E., & Ylvisåker, E. (2015). Reliability of objectively measured sedentary time and physical activity in adults. *PLoS One, 10*(7), Article e0133296. https://doi.org/10.1371/journal.pone.0133296

Banks, M.A. (2020). Tech giants, armed with wearables data, are entrenching in health research. *Nature Medicine, 26*(1), 4–5. https://doi.org/10.1038/s41591-019-0701-2

Bassett, D.R., Freedson, P.S., & John, D. (2019). Wearable activity trackers in clinical research and practice. *Kinesiology Review, 8*(1), 11–15. https://doi.org/10.1123/kr.2018-0060

Bergman, P. (2018). The number of repeated observations needed to estimate the habitual physical activity of an individual to a given level of precision. *PLoS One, 13*(2), Article e0192117. https://doi.org/10.1371/journal.pone.0192117

Bietz, M.J., Bloss, C.S., Calvert, S., Godino, J.G., Gregory, J., Claffey, M.P., Sheehan, J., & Patrick, K. (2016). Opportunities and challenges in the use of personal health data for health research. *Journal of the American Medical Informatics Association, 23*(e1), e42–e48. https://doi.org/10.1093/jamia/ocv118

Brage, S., Assah, F., & Msyamboza, K.P. (2020). Quantifying population levels of physical activity in Africa using wearable sensors: Implications for global physical activity surveillance. *BMJ Open Sport & Exercise Medicine, 6*(1), Article e000941. https://doi.org/10.1136/bmjsem-2020-000941

Brage, S., Ekelund, U., Brage, N., Hennings, M.A., Froberg, K., Franks, P.W., & Wareham, N.J. (2007). Hierarchy of individual calibration levels for heart rate and accelerometry to measure physical activity. *Journal of Applied Physiology, 103*(2), 682–692. https://doi.org/10.1152/japplphysiol.00092.2006

Brage, S., Westgate, K., Franks, P.W., Stegle, O., Wright, A., Ekelund, U., & Wareham, N.J. (2015). Estimation of free-living energy expenditure by heart rate and movement sensing: A doubly-labelled water study. *PLoS One, 10*(9), Article e0137206. https://doi.org/10.1371/journal.pone.0137206

Cho, S., Ensari, I., Weng, C., Kahn, M.G., & Natarajan, K. (2021). Factors affecting the quality of person-generated wearable device data and associated challenges: Rapid systematic review. *JMIR mHealth and uHealth, 9*(3), Article e20738. https://doi.org/10.2196/20738

Cosoli, G., Spinsante, S., & Scalise, L. (2020). Wrist-worn and chest-strap wearable devices: Systematic review on accuracy and metrological characteristics. *Measurement, 159,* Article 107789. https://doi.org/10.1016/j.measurement.2020.107789

Cui, Y., Chipchase, J., & Ichikawa, F. (2007). A cross culture study on phone carrying and physical personalization. In N. Aykin (Ed.), *Usability and internationalization. HCI and culture* (Vol. 4559, pp. 483–492). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-73287-7_57

Doherty, A., Jackson, D., Hammerla, N., Plötz, T., Olivier, P., Granat, M.H., White, T., van Hees, V.T., Trenell, M.I., Owen, C.G., Preece, S.J., Gillions, R., Sheard, S., Peakman, T., Brage, S., & Wareham, N.J. (2017). Large scale population assessment of physical activity using wrist worn accelerometers: The UK biobank study. *PLoS One, 12*(2), Article e0169649. https://doi.org/10.1371/journal.pone.0169649

Düking, P., Fuss, F.K., Holmberg, H.-C., & Sperlich, B. (2018). Recommendations for assessment of the reliability, sensitivity, and validity of data provided by wearable sensors designed for monitoring physical activity. *JMIR mHealth and uHealth, 6*(4), e102. https://doi.org/10.2196/mhealth.9341

Endeavour Partners. (2014). Inside Wearables Part 1: How behavior change unlocks long-term engagement. https://medium.com/@endeavourprtnrs/inside-wearable-how-the-science-of-human-behavior-change-offers-the-secret-to-long-term-engagement-a15b3c7d4cf3

Ernsting, C., Dombrowski, S.U., Oedekoven, M., O´Sullivan, J.L., Kanzler, M., Kuhlmey, A., & Gellert, P. (2017). Using smartphones and health apps to change and manage health behaviors: A population-based survey. *Journal of Medical Internet Research, 19*(4), e101. https://doi.org/10.2196/jmir.6838

Evenson, K.R., Goto, M.M., & Furberg, R.D. (2015). Systematic review of the validity and reliability of consumer-wearable activity trackers. *International Journal of Behavioral Nutrition and Physical Activity, 12*(1), 159. https://doi.org/10.1186/s12966-015-0314-1

Feehan, L.M., Geldman, J., Sayre, E.C., Park, C., Ezzat, A.M., Yoo, J.Y., Hamilton, C.B., & Li, L.C. (2018). Accuracy of Fitbit devices: Systematic review and narrative syntheses of quantitative data. *JMIR mHealth and uHealth, 6*(8), Article e10527. https://doi.org/10.2196/10527

Ferguson, T., Rowlands, A.V., Olds, T., & Maher, C. (2015). The validity of consumer-level, activity monitors in healthy adults worn in free-

Brought to you by CAMBRIDGE UNIVERSITY | Unauthenticated | Downloaded 02/04/22 10:31 AM UTC

living conditions: A cross-sectional study. *International Journal of Behavioral Nutrition and Physical Activity, 12*(1), 42. https://doi.org/10.1186/s12966-015-0201-9

Fuller, D., Colwell, E., Low, J., Orychock, K., Tobin, M.A., Simango, B., Buote, R., Van Heerden, D., Luan, H., Cullen, K., Slade, L., & Taylor, N.G.A. (2020). Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: Systematic review. *JMIR mHealth and uHealth, 8*(9), Article e18694. https://doi.org/10.2196/18694

Fulton, J.E., Carlson, S.A., Ainsworth, B.E., Berrigan, D., Carlson, C., Dorn, J.M., Heath, G.W., Kohl, H.W., Lee, I.-M., Lee, S.M., Másse, L.C., Morrow, J.R., Gabriel, K.P., Pivarnik, J.M., Pronk, N.P., Rodgers, A.B., Saelens, B.E., Sallis, J.F., Troiano, R.P., . . . Wendel, A. (2016). Strategic priorities for physical activity surveillance in the United States. *Medicine & Science in Sports & Exercise, 48*(10), 2057–2069. https://doi.org/10.1249/MSS.0000000000000989

Gray, L., McCartney, G., White, I.R., Katikireddi, S.V., Rutherford, L., Gorman, E., & Leyland, A.H. (2013). Use of record-linkage to handle non-response and improve alcohol consumption estimates in health survey data: A study protocol. *BMJ Open, 3*(3), Article e002647. https://doi.org/10.1136/bmjopen-2013-002647

Hendker, A., Jetzke, M., Eils, E., & Voelcker-Rehage, C. (2020). The implication of wearables and the factors affecting their usage among recreationally active people. *International Journal of Environmental Research and Public Health, 17*(22), Article 8532. https://doi.org/10.3390/ijerph17228532

Henriksen, A., Haugen Mikalsen, M., Woldaregay, A.Z., Muzny, M., Hartvigsen, G., Hopstock, L.A., & Grimsgaard, S. (2018). Using fitness trackers and smartwatches to measure physical activity in research: Analysis of consumer wrist-worn wearables. *Journal of Medical Internet Research, 20*(3), e110. https://doi.org/10.2196/jmir.9157

Hicks, J.L., Althoff, T., Sosic, R., Kuhar, P., Bostjancic, B., King, A.C., Leskovec, J., & Delp, S.L. (2019). Best practices for analyzing large-scale health data from wearables and smartphone apps. *npj Digital Medicine, 2*(1), 45. https://doi.org/10.1038/s41746-019-0121-1

Hyde, E.T., Omura, J.D., Fulton, J.E., Weldy, A., & Carlson, S.A. (2020). Physical activity surveillance using wearable activity monitors: Are US adults willing to share their data? *American Journal of Health Promotion, 34*(6), 672–676. https://doi.org/10.1177/0890117119900587

Laranjo, L., Ding, D., Heleno, B., Kocaballi, B., Quiroz, J.C., Tong, H.L., Chahwan, B., Neves, A.L., Gabarron, E., Dao, K.P., Rodrigues, D., Neves, G.C., Antunes, M.L., Coiera, E., & Bates, D.W. (2021). Do smartphone applications and activity trackers increase physical activity in adults? Systematic review, meta-analysis and metaregression. *British Journal of Sports Medicine, 55*(8), 422–432. https://doi.org/10.1136/bjsports-2020-102892

Mair, J.L., Hayes, L.D., Campbell, A.k., & Sculthorpe, N. (2021). Should we use activity tracker data from smartphones and wearables to understand population physical activity patterns? *Journal for the Measurement of Physical Behaviour*. https://doi.org/10.1123/jmpb.2021-0012

Mercer, K., Li, M., Giangregorio, L., Burns, C., & Grindrod, K. (2016). Behavior change techniques present in wearable activity trackers: A critical analysis. *JMIR mHealth and uHealth, 4*(2), e40. https://doi.org/10.2196/mhealth.4461

Meyer, J., Wasmann, M., Heuten, W., El Ali, A., & Boll, S.C.J. (2017). Identification and classification of usage patterns in long-term activity tracking. Paper presented at the Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 667–678. https://doi.org/10.1145/3025453.3025690

Omura, J.D., Carlson, S.A., Paul, P., Watson, K.B., & Fulton, J.E. (2017). National physical activity surveillance: Users of wearable activity monitors as a potential data source. *Preventive Medicine Reports, 5,* 124–126. https://doi.org/10.1016/j.pmedr.2016.10.014

Omura, J.D., Whitfield, G.P., Chen, T.J., Hyde, E.T., Ussery, E.N., Watson, K.B., & Carlson, S.A. (2021). Surveillance of physical activity and sedentary behavior among youth and adults in the United States: History and opportunities. *Journal of Physical Activity and Health, 18*(Suppl. 1), S6–S24. https://doi.org/10.1123/jpah.2021-0179

Patel, M.S., Asch, D.A., & Volpp, K.G. (2015). Wearable devices as facilitators, not drivers, of health behavior change. *Journal of the American Medical Association, 313*(5), 459–460. https://doi.org/10.1001/jama.2014.14781

Pearce, M., Bishop, T.R.P., Sharp, S., Westgate, K., Venables, M., Wareham, N.J., & Brage, S. (2020). Network harmonization of physical activity variables through indirect validation. *Journal for the Measurement of Physical Behaviour, 3*(1), 8–18. https://doi.org/10.1123/jmpb.2019-0001

Pontin, F., Lomax, N., Clarke, G., & Morris, M.A. (2021). Socio-demographic determinants of physical activity and app usage from smartphone data. *Social Science & Medicine, 284,* Article 114235. https://doi.org/10.1016/j.socscimed.2021.114235

Redmayne, M. (2017). Where's your phone? A survey of where women aged 15-40 carry their smartphone and related risk perception: A survey and pilot study. *PLoS One, 12*(1), Article e0167996. https://doi.org/10.1371/journal.pone.0167996

Richiardi, L., Pizzi, C., & Pearce, N. (2013). Commentary: Representativeness is usually not necessary and often should be avoided. *International Journal of Epidemiology, 42*(4), 1018–1022. https://doi.org/10.1093/ije/dyt103

Ried-Larsen, M., Brønd, J.C., Brage, S., Hansen, B.H., Grydeland, M., Andersen, L.B., & Møller, N.C. (2012). Mechanical and free living comparisons of four generations of the Actigraph activity monitor. *International Journal of Behavioral Nutrition and Physical Activity, 9*(1), 113. https://doi.org/10.1186/1479-5868-9-113

Rising, C.J., Jensen, R.E., Moser, R.P., & Oh, A. (2020). Characterizing the US population by patterns of mobile health use for health and behavioral tracking: Analysis of the National Cancer Institute's Health Information National Trends Survey Data. *Journal of Medical Internet Research, 22*(5), Article e16299. https://doi.org/10.2196/16299

Rothman, K.J., Gallacher, J.E., & Hatch, E.E. (2013). Why representativeness should be avoided. *International Journal of Epidemiology, 42*(4), 1012–1014. https://doi.org/10.1093/ije/dys223

Rowlands, A.V., Dawkins, N.P., Maylor, B., Edwardson, C.L., Fairclough, S.J., Davies, M.J., Harrington, D.M., Khunti, K., & Yates, T. (2019). Enhancing the value of accelerometer-assessed physical activity: Meaningful visual comparisons of data-driven translational accelerometer metrics. *Sports Medicine—Open, 5*(1), 47. https://doi.org/10.1186/s40798-019-0225-9

Statista. (2021). Wearables—Statistics & facts. https://www.statista.com/topics/1556/wearable-technology/

Strain, T., Sharp, S.J., Spiers, A., Price, H., Williams, C., Fraser, C., Brage, S., Wijndaele, K., & Kelly, P. (2022). Population level physical activity before and during the first national COVID-19 lockdown: A nationally representative repeat cross-sectional study of 5 years of Active Lives data in England. *The Lancet Regional Health—Europe, 12,* Article 100265. https://doi.org/10.1016/j.lanepe.2021.100265

Strain, T., Wijndaele, K., & Brage, S. (2019). Physical activity surveillance through smartphone apps and wearable trackers: Examining the UK potential for nationally representative sampling. *JMIR mHealth and uHealth, 7*(1), Article e11898. https://doi.org/10.2196/11898

Brought to you by CAMBRIDGE UNIVERSITY | Unauthenticated | Downloaded 02/04/22 10:31 AM UTC

Strath, S.J., & Rowley, T.W. (2018). Wearables for promoting physical activity. *Clinical Chemistry, 64*(1), 53–63. https://doi.org/10.1373/clinchem.2017.272369

Swan, M. (2013). The quantified self: Fundamental disruption in big data science and biological discovery. *Big Data, 1*(2), 85–99. https://doi.org/10.1089/big.2012.0002

Toepoel, V., Luiten, A., & Zandvliet, R. (2021). Response, willingness, and data donation in a study on accelerometer possession in the general population. *Survey Practice, 14*(1), 1–19. https://doi.org/10.29115/SP-2021-0005

Turner, K., Jo, A., Wei, G., Tabriz, A.A., Clary, A., & Jim, H.S.L. (2021). Sharing patient-generated data with healthcare providers: Findings from a 2019 national survey. *Journal of the American Medical Informatics Association, 28*(2), 371–376. https://doi.org/10.1093/jamia/ocaa272

van Hees, V.T., Fang, Z., Langford, J., Assah, F., Mohammad, A., da Silva, I.C.M., Trenell, M.I., White, T., Wareham, N.J., & Brage, S. (2014). Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: An evaluation on four continents. *Journal of Applied Physiology, 117*(7), 738–744. https://doi.org/10.1152/japplphysiol.00421.2014

Wareham, N.J., Wong, M.-Y., Hennings, S., Mitchell, J., Rennie, K., Cruickshank, K., & Day, N.E. (2000). Quantifying the association between habitual energy expenditure and blood pressure. *International Journal of Epidemiology, 29*(4), 655–660. https://doi.org/10.1093/ije/29.4.655

Welk, G.J., Bai, Y., Lee, J.-M., Godino, J., Saint-Maurice, P.F., & Carr, L. (2019). Standardizing analytic methods and reporting in activity monitor validation studies. *Medicine & Science in Sports & Exercise, 51*(8), 1767–1780. https://doi.org/10.1249/MSS.0000000000001966

White, T., Westgate, K., Hollidge, S., Venables, M., Olivier, P., Wareham, N., & Brage, S. (2019). Estimating energy expenditure from wrist and thigh accelerometry in free-living adults: A doubly labelled water study. *International Journal of Obesity, 43*(11), 2333–2342. https://doi.org/10.1038/s41366-019-0352-x

White, T., Westgate, K., Wareham, N.J., & Brage, S. (2016). Estimation of physical activity energy expenditure during free-living from wrist accelerometry in UK adults. *PLoS One, 11*(12), Article e0167472. https://doi.org/10.1371/journal.pone.0167472

Wright, S.P., Hall Brown, T.S., Collier, S.R., & Sandberg, K. (2017). How consumer physical activity monitors could transform human physiology research. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 312*(3), R358–R367. https://doi.org/10.1152/ajpregu.00349.2016

Xie, Z., Jo, A., & Hong, Y.-R. (2020). Electronic wearable device and physical activity among US adults: An analysis of 2019 HINTS data. *International Journal of Medical Informatics, 144,* Article 104297. https://doi.org/10.1016/j.ijmedinf.2020.104297

Brought to you by CAMBRIDGE UNIVERSITY | Unauthenticated | Downloaded 02/04/22 10:31 AM UTC