American Journal of
Preventive Medicine

RESEARCH LETTER

Assessing Validity of the Fitbit Indicators for U.S. Public Health Surveillance

Kelly R. Evenson, PhD, MS,[1] Fang Wen, MS, MCS,[1] Robert D. Furberg, PhD, MBA[2]

INTRODUCTION

Personally generated health data are increasingly used to report on population prevalence and trends, providing a new avenue for public health surveillance.[1] Documentation of acceptable measurement properties to ensure correct interpretations should precede their use. One common source of personally generated health data comes from activity trackers, self-worn devices that provide feedback and long-term tracking on physical activity–related metrics.[2] Activity trackers are relatively unobtrusive and low cost, with 12.5% of U.S. adults reporting wearing one in 2015.[3] Companies selling activity trackers already report on data acquired by their users.[4,5]

In 2015, the U.S. Fitbit Health and Activity Index™ was launched (and updated in 2017), providing a suite of metrics including (1) prevalence of five indicators (steps, active minutes, resting heart rate, sleep, BMI), (2) popular Fitbit activities, and (3) time trends in activities. Using company-provided online tools, users can cross-tabulate three Fitbit indicators (steps, active minutes, resting heart rate) with diabetes, obesity, or cardiovascular disease (from the 2014 Behavioral Risk Factor Surveillance System [BRFSS]). An expert panel recommended assessing the psychometric properties of instruments for surveillance,[6] but the validity of these Fitbit indicators is unknown. Thus, this study explored whether the Fitbit indicators of physical activity (steps, active minutes), resting heart rate, and BMI provided evidence for validity for use as a surveillance tool.

METHODS

The Fitbit company evaluated aggregated data from >10 million users between June 2015 and June 2016 and published results in 2017. In February 2017, average steps/day, active minutes/day, resting heart rate, and BMI were abstracted by state or district from their website (www.fitbit.com/activity-index). All measures except BMI were Fitbit-assessed. Height and weight were entered typically at account set up.

These data were compared to state- or district-based data from the 2015 BRFSS (www.cdc.gov/brfss/). The BRFSS is an ongoing, state-based random-digit dialed telephone survey of noninstitutionalized adults aged $\geq$18 years. Participants self-reported about physical activity or exercise in the past month, including the type, duration, and frequency of up to two activities. Physical activities were summed in minutes/week for both total and vigorous intensity.[7] Estimated maximal oxygen uptake (VO_2) was age–gender specific.[7] BMI was derived in kg/m^2 using self-reported height and weight.

Spearman rank correlation coefficients provided associations between BRFSS and Fitbit indicators. As a guide, these ratings indicated agreement level[8]: 0–0.2 poor, 0.2–0.4 fair, 0.4–0.6 moderate, 0.6–0.8 substantial, and 0.8–less than 1.0 almost perfect. Bland–Altman plot for BMI indicated direction of bias.[9] Analyses were conducted using SAS, version 9.3, and data from both sources were deidentified and publicly available.

RESULTS

Both steps and active minutes Fitbit indicators showed a poor association with VO_2 and a fair association with vigorous activity (Table 1). The resting heart rate Fitbit indicator showed a poor association with VO_2 and total physical activity, and a fair association with vigorous activity. The BMI Fitbit indicator showed a fair association with BMI.

DISCUSSION

This study found correlations postulated to be associated with four Fitbit indicators were poor or fair in strength, indicating concerns with using these data as state-based indicators. However, it is encouraging that correlations with Fitbit steps, active minutes, and resting heart rate were stronger for vigorous activity, which is usually better recalled compared to total activity, indicating

From the [1]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina–Chapel Hill, Chapel Hill, North Carolina; and [2]Digital Health and Clinical Informatics, RTI International, Research Triangle Park, North Carolina

Address correspondence to: Kelly R. Evenson, PhD, MS, 137 East Franklin Street, Suite 306, University of North Carolina, Gillings School of Global Public Health, Department of Epidemiology, Chapel Hill NC 27514. E-mail: kelly_evenson@unc.edu.

0749-3797/$36.00
https://doi.org/10.1016/j.amepre.2017.06.005

Table 1. Spearman's Rank Correlation Coefficients Between Fitbit Indicators and BRFSS Measures[a]

From 2015 BRFSS	Fitbit BMI	Fitbit Steps/day	Fitbit Active minutes/day	Fitbit Resting heart rate/day
BMI	0.25[b]	−0.24	**−0.32**	**0.56**
Maximal oxygen uptake, (milliliters/kilogram/minute)*100	−0.08	−0.14	−0.04	−0.04
Total physical activity, minutes/week	−0.07	0.15	0.11	−0.14
Vigorous physical activity, minutes/week	−0.12	0.21	0.20	**−0.31**

Note: Boldface indicates statistical significance ($p < 0.05$) from rho=0; all other correlations have $p \geq 0.05$.
[a]All measures in the table represent averages at the state level (50 states and Washington, DC; N=51). Outliers from the BRFSS data were removed before calculating the weighted average for each state/district. Outliers were defined as <1st and >99th percentile for BMI and resting heart rate, and >99th percentile for maximal oxygen uptake and physical activity. The BRFSS survey weight calculation is explained elsewhere (www.cdc.gov/brfss/annual_data/2015/pdf/weighting_the-data_webpage_content.pdf).
[b]The average of the difference in BMI from the Bland Altman plot was 0.18 and the limit of agreement was −0.85 and 1.21, indicating that on average the Fitbit BMI measured 0.18 kg/m^2 more than the BRFSS BMI.
BRFSS, Behavioral Risk Factor Surveillance System.

some specificity. A 2015 national survey reported that activity tracker users are not representative of the U.S. adult population.[3] Based on the website documentation, the Fitbit indicators do not seem to be weighted to any population, thus contributing to these low correlations,[1] in addition to measurement (self-report versus directly assessed) differences.

Limitations

This study has several limitations. The BRFSS data are self-reported, thus subject to social desirability and recall biases, and vary in terms of validity and reliability.[10] CIs are not provided due to the reporting of the Fitbit data, and documentation on data cleaning was not available. The two data sources only partially aligned temporally (2015–2016 Fitbit data versus 2015 BRFSS).

CONCLUSIONS

This study revealed that the Fitbit indicators did not correlate well with state- or district-based indicators. Technology companies continue extending available features of wearable devices, improving data processing algorithms, and enhancing individualized feedback. Although enthusiasm for the use of such data for public health surveillance and interventions increases, companies are encouraged to derive metrics that are valid, reliable, and generalizable.

ACKNOWLEDGMENTS

We gratefully acknowledge funding provided by the North Carolina Translational and Clinical Sciences Institute (NIH grant #UL1TR001111). The content is solely the responsibility of the authors and does not necessarily represent the official views of NIH or RTI International.

No financial disclosures were reported by the authors of this paper.

REFERENCES

1. Chunara R, Wisk LE, Weitzman ER. Denominator issues for personally generated data in population health monitoring. *Am J Prev Med*. 2017;52(4):549–553. https://doi.org/10.1016/j.amepre.2016.10.038.
2. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. *Int J Behav Nutr Phys Act*. 2015;12:159. https://doi.org/10.1186/s12966-015-0314-1.
3. Omura JD, Carlson SA, Paul P, et al. National physical activity surveillance: users of wearable activity monitors as a potential data source. *Prev Med Rep*. 2016;5:124–126. https://doi.org/10.1016/j.pmedr.2016.10.014.
4. Fitbit Inc. Weathering the weather. www.fitbit.com/weathermap. Accessed April 19, 2017.
5. Mohan S. The Jawbone Blog: What makes people happy? We have the data. https://jawbone.com/blog/what-makes-people-happy/. Published June 4, 2015. Accessed April 19, 2017.
6. Fulton JE, Carlson SA, Ainsworth BE, et al. Strategic priorities for physical activity surveillance in the United States. *Med Sci Sports Exerc*. 2016;48(10):2057–2069. https://doi.org/10.1249/MSS.0000000000000989.
7. CDC. A Data Users Guide to the BRFSS Physical Activity Questions. Atlanta, GA: U.S. DHHS, CDC. www.cdc.gov/brfss/pdf/PA%20RotatingCore_BRFSSGuide_508Comp_07252013FINAL.pdf. Accessed April 19, 2017.
8. Landis J, Koch G. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174. https://doi.org/10.2307/2529310.
9. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;327(8476):307–310. https://doi.org/10.1016/S0140-6736(86)90837-8.
10. Pierannunzi C, Hu SS, Balluz L. A systematic review of publications assessing reliability and validity of the Behavioral Risk Factor Surveillance System (BRFSS), 2004-2011. *BMC Med Res Methodol*. 2013;13:49. https://doi.org/10.1186/1471-2288-13-49.