**VALIDITY OF COMMERCIALLY AVAILABLE WEARABLE DEVICES FOR MEASURING STEP COUNT BY AGE, GENDER, BMI AND WEAR LOCATION**

**HKR 6000 FINAL MAJOR ASSIGNMENT**

**SUBMITTED BY**

**SUMAYYAH MUSA**

**APRIL 13, 2021**

**TO BE SUBMITTED IN THE JOURNAL OF MEDICAL INTERNET RESEARCH MHEALTH AND UHEALTH**

**ABSTRACT**

**Background**: Consumer wearable devices are electronic devices used for monitoring fitness and other health-related metrics. They provide an accessible means to objectively measure physical activity through step counts.

**Objective**: The purpose of this project is to examine the validity of consumer wearables in measuring step count in target populations.

**Methods**: Data taken from a recent systematic review on the reliability and validity of wearables by Fuller (2020) was quantitatively analyzed for the validity of step count in different testing conditions and population groups

**Results**: The devices were generally accurate in measuring step count in the free-living conditions in all population.

**Conclusions**: Fitbit and Garmin devices were generally the most accurate regardless of testing condition. Age and BMI have the greatest influence in the population group examined on validity. Future research should be done with more sample size.

**Keywords:** PA, Wearable devices, step count, testing conditions, Fitbit, Apple,

**INTRODUCTION**

Physical inactivity has been identified as the fourth leading risk factor for global mortality, and its levels are rising in many countries with major implications for the prevalence of noncommunicable diseases (NCDs) and the general health of the population worldwide. A recent report suggested that more than 80% of the world's adolescent population is insufficiently physically active (WHO, 2020). Government organizations have attempted to improve these numbers by implementing initiatives aimed at promoting physical activity. Though the successful promotion of physical activity is a complex multi-facetted issue, behavior change is a well-established method to increase physical activity (Conn et al, 2011). One of the most common measures of PA levels provided by these consumer devices is step count, and metrics from commercial wearable devices have been developed, including 10,000 steps per day (Tudor-Locke et al, 2011) and 100 steps per minute for moderate to vigorous activity. However, research has shown variation in step count among devices, and the applicability of these metrics maybe significantly influenced by age (Boyer et al, 2017), BMI (Molina-Garcia et al, 2019), sex (Telfer & Bigham, 2019), testing conditions (O’Connell et al, 2017) and wear location (Block et al, 2019).

Step count is easily understandable by the general population and there is a growing body of evidence supporting the use of these technological approaches in ‘passive’ PA monitoring and ‘active’ intervention (Johnson et al, 2020). The 2018 PA Guidelines Advisory Committee Report highlighted that step counting is an accessible means to monitor and set PA goals, and that recent evidence supports an inverse dose–response relationship between daily step count and all-cause mortality, cardiovascular events and type 2 diabetes (Hansen et al, 2020). Furthermore, ‘active’ interventions using wearable devices as a measurement tool can result in significant increases in PA participation, highlighting their potential utility in personalised medicine, increasing adherence to PA and to embed sustained healthy lifestyle habits (Brickwood et al, 2019).

In a recent study of 158 publications, Fuller et al, (2019) noted that in laboratory-based settings, Fitbit, Apple Watch, and Samsung appeared to measure steps accurately. How valid are these devices when it comes to population level measurement of physical activity? The purpose of this paper was to examine the accuracy of the wearable devices in measuring step count in target population. The information summarized herein can be used to inform specific consumers and can aid researchers in study design when selecting step count monitoring devices.

**METHODS**

**Design**

This study is a descriptive quantitative design which seeks to describe and compare the differences in the criterion validity of the wearables based on age, sex, BMI, testing conditions and wear location. The research question to be answered will be “Does variability in gender, age, BMI, and part of body on which the device is worn affect the validity of wearable devices in measuring step count?”.

**Data**

The publicly available dataset used in this paper was from the systematic review carried out by Fuller (2019) on the validity of consumer wearables, and was downloaded from the BeapLab Dataverse website (Harvard Dataverse, 2019). An in-depth web search of the available consumer-wearable models and their specifications (placement, size, weight, cost, and connectivity) first conducted and documented. The data extraction process then consisted of the following: (1) categorizing the selected full-text articles into reliability or validity studies (EC, JL, and DF); (2) using a modification of the modified Consensus-Based Standards for the Selection of Health Status Measurement Instruments (COSMIN) validation subscale used by Feehan et al [[13](https://mhealth.jmir.org/2020/9/e18694#ref13)] and an a priori modified COSMIN reliability subscale (to assess the quality and risk of bias of each study (EC and DF); (3) extracting the key characteristics from each selected publication and compiling them into tables. Details from each reviewer were compared, and inconsistencies were resolved through consensus before compiling the results (EC and DF).

Data extracted included characteristics of studies, participants, and devices, including study setting and activity type, outcomes measured, and type of criterion measure used. Correlation coefficients were extracted for all reliability comparisons reported in each study. Correlation coefficients, percentage difference and group mean values, MAPE values, and level-of-agreement data were extracted for all validity comparisons where available. Where group percentage differences were not reported, we calculated group percentage error ([wearablemean – criterionmean]/criterionmean × 100) to allow for comparison across studies. We split a small number of studies (n=10) into “substudies” (n=21), where separate populations were examined in the same publication (Fuller et al, 2020).

**Analysis Plan**

I explored a narrative synthesis of the available quantitative data within step count using group percentage difference (MPE) as the common metrics for criterion validity. My interpretation of measurement accuracy was focused on acceptable limits of percentage difference of ±3% in controlled settings and percentage difference of ±10% in free-living settings, as outlined in previous work (Feehan, 2018).

All quantitative analyses and plots will be done using RStudio version 1.4.1106 (RStudio Inc) and R version 4.0.4 (The R Foundation). The dataset will be imported into RStudio, cleaned and analyzed.

The variables that are necessary for the analysis will be checked for missing values. This includes population of male, population of female, mean age, the part of the body the device was worn, the mean percentage error, the study setting, and the outcome measured i.e. step count.

Age: The mean age of the participants in each study will be used to categorized them into groups as categorized by the WHO for age-recommended guidelines for Physical Activity:

* Children (5 – 17)
* Adults (18 – 64)
* Older adults (65 and above)

Sex: The percentage of males and females in a particular study will be determined, and the gender with the higher percentage will be the labelled as the study’s main gender. The studies with the same percentage of male and female percentages will be categorized as Neutral.

BMI: The mean BMI will be determined, and then categorized as follows:

* Normal Weight (<18.5)
* Overweight (18.5 – 24.9)
* Obese (30 & over)

Wear Location: The part of the body where the device was worn on for the activity will be determined. The devices were worn on the wrist, waist, chest, thigh, upper arm, and calf.

Setting: The condition where the testing took place, it could be in a laboratory controlled setting or a free living setting.

MPE: The summary statistics for the overall MPE will be determined, missing values and outliers will be checked for and removed or replaced depending on the variable under which they present. The interpretation of measurement accuracy will be focused on acceptable limits of percentage difference of ±3% in controlled settings and percentage difference of ±10% in free-living settings, as outlined in previous work (Feehan et al, 2018).

Boxplots will be used to check if the devices under, over or accurately estimated the step count for the different populations. Kruskal Wallis Test will be used to compare the differences in the validity for the groups because the data is unpaired, non-parametric and two-tailed.

Multiple pairwise comparisons will be used to compare all groups to determine the pair of groups that are different from each other. Dunn’s Test will be used since we don’t need to account for multiple testing. The Bonferroni method will be used because we don’t have equal sample size in each group, and the group variances are not similar.

Lastly, a one-way ANOVA will be carried out using sex, age, gender and wear location as factor to determine which variables influenced MPE the most.

**RESULTS**

**Descriptive Statistics of Participants (Table 1)**

Of the 747 publications that examined validity of step count, 602 (81%) of them examined validity in adults, 696 (92%) of them examined validity in laboratory-controlled setting. Two hundred and fifteen (29%) examined validity in normal weighted individuals (Mean 0.004%), while 228 (31%) did not take BMI into consideration (Mean -0.06%). Only seven brands measured step count and had more than 10 comparisons: Apple, Fitbit, Garmin, Misfit, Polar, Samsung and Withings.

**Table 1: DESCRIPTIVE STATISTICS OF VALIDITY FOR TARGET POPULATIONS (n = 747)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | **Frequency** | **Percent (%)** | **Mean MPE (%)** | **SD MPE** |
| **Age** | Children | 23 | 3.1 | 0.015 | 0.107 |
| Adults | 602 | 80.6 | -0.026 | 0.097 |
| Older Adults | 132 | 17.7 | -0.073 | 0.114 |
| **Sex** | Female | 193 | 25.8 | -0.019 | 0.097 |
| Male | 339 | 45.4 | -0.044 | 0.112 |
| Neutral | 225 | 30.1 | -0.027 | 0.088 |
| **BMI** | Normal weight | 215 | 28.8 | 0.004 | 0.089 |
| Overweight | 160 | 21.4 | -0.052 | 0.099 |
| Obese | 144 | 19.3 | -0.029 | 0.066 |
| NA | 228 | 30.5 | -0.055 | 0.123 |
| **Wear Location** | Wrist | 401 | 53.7 | -0.034 | 0.104 |
| Waist/Hip | 233 | 31.2 | -0.027 | 0.096 |
| Torso | 65 | 8.7 | -0.034 | 0.108 |
| Calf | 51 | 6.8 | -0.042 | 0.109 |
| Upper Arm | 6 | 0.8 | -0.021 | 0.020 |
| Thigh | 1 | 0.1 | -0.308 | NA |
| **Setting** | Controlled | 686 | 92 | -0.307 | 0.095 |
| Free | 61 | 8 | 0.021 | 0.157 |

**Figure 1: Mean Percentage Error (MPE) plots by study sample size for step count**

I explored the potential for bias related to sample size in step count by examining the percentage error dispersion in controlled (A) and free-living (B) settings by sample size using scatter plots.



In these examinations, I saw no apparent systematic bias for measurement error.

**Inferential Statistics**

**Validity: Controlled Settings**

I examined criterion validity for step count separately for controlled and free-living settings. For controlled settings, there was sufficient data to examine validity by age, sex, BMI, wear location and within devices.

**Validity for Step Count in Controlled Settings**

Group measurement error in controlled settings was reported for 680 out of the 747 studies regardless of the criterion measure. Of these, 52% (n=355) were within ±3% measurement error, 37% (n=249) were below −3% measurement error, and 11% (n=76) were above 3% measurement error. The overall tendency was to underestimate step count (mean: −0.04%, median: −0.01%).

**Validity in Controlled Settings by Brand and Target Population**

Figure 2 shows the mean percentage error of step count by device brand for age (A), sex (B), BMI (C) and wear location (D) groups with more than 10 comparisons.



**Validity of Step count by Brand and Age (Figure 2)**

Children: Only Fitbit had more than 10 comparisons, and it mostly underestimated step count.

Adults: Fitbit, Garmin and Withings accurately estimated step count, Misfit and fewer studies of Samsung overestimated step count, while Polar and few studies of Samsung and fewer studies of Apple underestimated step count.

Older Adults: Fitbit and fewer studies of Garmin underestimated step count, while more studies on Withings accurately measured step count.

**Validity of Step count by Brand and Gender (Figure 2)**

Female: All devices underestimated step count in studies with more females than males, except for Garmin and Withings that accurately estimated step count.

Male: Only Apple, Fitbit, Garmin and Withings accurately measured for step count in studies with more males than females.

Neutral: In studies with the same number of male and female participants, only Fitbit and Garmin accurately estimated step count.

**Validity of Step count by Brand and BMI (Figure 2)**

Normal weight: All devices were accurate except Misfit which overestimated step count.

Overweight: All devices were relatively accurate except Polar and Samsung which underestimated step count.

Obese: Except from Polar which underestimated step count, all other devices accurately measured step count.

NA: Some studies did not measure BMI, Apple and Samsung underestimated for these groups, while Fitbit, Garmin and Withings were accurate.

**Validity of Step count by Brand and Wear Location (Figure 2)**

Most of the devices were worn on the wrist for step count, and they were relatively accurate.

**Validity for Step Count in Free-Living Settings**

Group measurement error was reported for 47 out of the 747 studies regardless of the criterion measure. Of these, 51% (n=24) were within ±10% measurement error, 21% (n=10) were below −10% measurement error, and 28% (n=13) were above 10% measurement error. The overall tendency was to accurately estimate step count (mean: 0.001%, median: 0.02%).

**Validity in Controlled Settings by Brand and Target Population**

Figure 3 shows the mean percentage error of step count by device brand for age (A), sex (B), BMI (C) and wear location (D) groups in free-living settings, with more than 10 comparisons.



**Validity of Step count by Brand and Age (Figure 3)**

Adults: All devices generally measured step count accurately.

Older Adults: Misfit was accurate while Fitbit over estimated step count.

**Validity of Step count by Brand and Gender (Figure 2)**

All devices accurately estimated step count in studies with more females than males, and those with more males than females.

**Validity of Step count by Brand and BMI (Figure 2)**

All devices were accurate except Fitbit which overestimated step count in normal and overweight groups.

**Validity of Step count by Brand and Wear Location (Figure 2)**

Wrist: Fitbit overestimated step count, Misfit underestimated step count and all other devices worn on the wrist for step count were relatively accurate.

Waist/Hip: The devices accurately measured step count.

**Comparing groups (Figure 4)**

Figure 4 shows that at p-value <0.05, there was a statistically significant difference in the MPE of step count in the following order:

1. Between adults and older adult (p = 8.73e-08)
2. Between normal weight and overweight group (p = 2.45e-08)
3. Between normal weight and obese group (p = 6.00e-04)
4. Between children and older adults (p = 2.02e-03)

After running a One-way Anova, Age and BMI had the greatest influence on the validity of wearable devices.



**DISCUSSION/CONCLUSIONS**

Generally, Fitbit and Garmin were the most accurate devices all over the population group. Future studies should focus on examining the validity of these devices in specific populations. Age and BMI have the greatest influence on validity in this study.

Subsequent studies should recruits as much participants as possible, in order to accurately determine the validity of these devices.

**REFERENCES**

World Health Organization. Physical Activity Fact Sheet. World Health Organization.   URL: <https://www.who.int/news-room/fact-sheets/detail/physical-activity>  [accessed 2021-04-13]

Conn VS, Hafdahl AR, Mehr DR. Interventions to increase physical activity among healthy adults: meta-analysis of outcomes. Am J Public Health 2011 Apr;101(4):751-758 [[FREE Full text](http://europepmc.org/abstract/MED/21330590)] [[CrossRef](https://dx.doi.org/10.2105/AJPH.2010.194381" \t "_blank)] [[Medline](https://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=21330590&dopt=Abstract)]

Tudor-Locke C, Craig CL, Brown WJ, Clemes SA, De CK, Giles-Corti B, et al. How many steps/day are enough? For adults. Int J Behav Nutr Phys Act 2011 Jul 28;8:79 [[FREE Full text](https://ijbnpa.biomedcentral.com/articles/10.1186/1479-5868-8-79)] [[CrossRef](https://dx.doi.org/10.1186/1479-5868-8-79" \t "_blank)] [[Medline](https://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=21798015&dopt=Abstract)]

O’Connell  S, ÓLaighin  G, Quinlan  LR. When a step is not a step! specificity analysis of five physical activity monitors. PLoS One 2017;**12**:e0169616. [doi:10.1371/journal.pone.0169616](http://dx.doi.org/10.1371/journal.pone.0169616) [Google Scholar](file:////lookup/google-scholar%3flink_type=googlescholar&gs_type=article&author%5b0%5d=S+O’Connell&author%5b1%5d=G+ÓLaighin&author%5b2%5d=LR+Quinlan&title=When+a+step+is+not+a+step!+specificity+analysis+of+five+physical+activity+monitors&publication_year=2017&journal=PLoS+One&volume=12)

Block  VJ, Zhao  C, Hollenbach  JA, *et al*. Validation of a consumer-grade activity monitor for continuous daily activity monitoring in individuals with multiple sclerosis. Mult Scler J Exp Transl Clin 2019;**5**:205521731988866. [doi:10.1177/2055217319888660](http://dx.doi.org/10.1177/2055217319888660) pmid: http://www.ncbi.nlm.nih.gov/pubmed/31803492 [PubMed](https://bjsm.bmj.com/lookup/external-ref?access_num=http://www.n&link_type=MED&atom=%2Fbjsports%2Fearly%2F2020%2F12%2F23%2Fbjsports-2020-103147.atom)[Google Scholar](https://bjsm.bmj.com/lookup/google-scholar?link_type=googlescholar&gs_type=article&author%5B0%5D=VJ+Block&author%5B1%5D=C+Zhao&author%5B2%5D=JA+Hollenbach&title=Validation+of+a+consumer-grade+activity+monitor+for+continuous+daily+activity+monitoring+in+individuals+with+multiple+sclerosis&publication_year=2019&journal=Mult+Scler+J+Exp+Transl+Clin&volume=5)

Molina-Garcia  P, Migueles  JH, Cadenas-Sanchez  C, *et al* . A systematic review on biomechanical characteristics of walking in children and adolescents with overweight/obesity: possible implications for the development of musculoskeletal disorders. Obes Rev 2019;**20**:1033–44.[doi:10.1111/obr.12848](http://dx.doi.org/10.1111/obr.12848)pmid:http://www.ncbi.nlm.nih.gov/pubmed/30942558 [PubMed](https://bjsm.bmj.com/lookup/external-ref?access_num=http://www.n&link_type=MED&atom=%2Fbjsports%2Fearly%2F2020%2F12%2F23%2Fbjsports-2020-103147.atom)[Google Scholar](https://bjsm.bmj.com/lookup/google-scholar?link_type=googlescholar&gs_type=article&author%5B0%5D=P+Molina-Garcia&author%5B1%5D=JH+Migueles&author%5B2%5D=C+Cadenas-Sanchez&title=A+systematic+review+on+biomechanical+characteristics+of+walking+in+children+and+adolescents+with+overweight/obesity:+possible+implications+for+the+development+of+musculoskeletal+disorders&publication_year=2019&journal=Obes+Rev&volume=20&pages=1033-44)

Boyer  KA, Johnson  RT, Banks  JJ, *et al*. Systematic review and meta-analysis of gait mechanics in young and older adults. Exp Gerontol 2017;**95**:63–70.[doi:10.1016/j.exger.2017.05.005](http://dx.doi.org/10.1016/j.exger.2017.05.005) [CrossRef](https://bjsm.bmj.com/lookup/external-ref?access_num=10.1016/j.exger.2017.05.005&link_type=DOI)[Google Scholar](https://bjsm.bmj.com/lookup/google-scholar?link_type=googlescholar&gs_type=article&author%5B0%5D=KA+Boyer&author%5B1%5D=RT+Johnson&author%5B2%5D=JJ+Banks&title=Systematic+review+and+meta-analysis+of+gait+mechanics+in+young+and+older+adults&publication_year=2017&journal=Exp+Gerontol&volume=95&pages=63-70)

Telfer  S, Bigham  JJ. The influence of population characteristics and measurement system on barefoot plantar pressures: a systematic review and meta-regression analysis. Gait Posture 2019;**67**:269–76. [doi:10.1016/j.gaitpost.2018.10.030](http://dx.doi.org/10.1016/j.gaitpost.2018.10.030) [Google Scholar](https://bjsm.bmj.com/lookup/google-scholar?link_type=googlescholar&gs_type=article&author%5B0%5D=S+Telfer&author%5B1%5D=JJ+Bigham&title=The+influence+of+population+characteristics+and+measurement+system+on+barefoot+plantar+pressures:+a+systematic+review+and+meta-regression+analysis&publication_year=2019&journal=Gait+Posture&volume=67&pages=269-76)

Hansen  BH, Dalene  KE, Ekelund  U, *et al*. Step by step: association of device-measured daily steps with all-cause mortality-A prospective cohort study. Scand J Med Sci Sports 2020;**30**:1705–11.[doi:10.1111/sms.13726](http://dx.doi.org/10.1111/sms.13726) pmid: <http://www.ncbi.nlm.nih.gov/pubmed/32427398> [PubMed](https://bjsm.bmj.com/lookup/external-ref?access_num=http://www.n&link_type=MED&atom=%2Fbjsports%2Fearly%2F2020%2F12%2F23%2Fbjsports-2020-103147.atom)[Google Scholar](https://bjsm.bmj.com/lookup/google-scholar?link_type=googlescholar&gs_type=article&author%5B0%5D=BH+Hansen&author%5B1%5D=KE+Dalene&author%5B2%5D=U+Ekelund&title=Step+by+step:+association+of+device-measured+daily+steps+with+all-cause+mortality-A+prospective+cohort+study&publication_year=2020&journal=Scand+J+Med+Sci+Sports&volume=30&pages=1705-11)

Brickwood  K-J, Watson  G, O'Brien  J, *et al*. Consumer-Based wearable activity Trackers increase physical activity participation: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019; **7**:e11819. [doi:10.2196/11819](http://dx.doi.org/10.2196/11819) [Google Scholar](https://bjsm.bmj.com/lookup/google-scholar?link_type=googlescholar&gs_type=article&author%5B0%5D=K-J+Brickwood&author%5B1%5D=G+Watson&author%5B2%5D=J+O'Brien&title=Consumer-Based+wearable+activity+Trackers+increase+physical+activity+participation:+systematic+review+and+meta-analysis&publication_year=2019&journal=JMIR+Mhealth+Uhealth&volume=7)

Johnston W, Judice PB, Molina García P*, et al.* Recommendations for determining the validity of consumer wearable and smartphone step count: expert statement and checklist of the INTERLIVE network. *British Journal of Sports Medicine*Published Online First: 24 December 2020.  doi:10.1136/bjsports-2020-103147

Fuller D, Colwell E, Low J, Orychock K, Tobin M, Simango B, Buote R, Van Heerden D, Luan H, Cullen K, Slade L, Taylor N. Reliability and Validity of Commercially Available Wearable Devices for Measuring Steps, Energy Expenditure, and Heart Rate: Systematic Review JMIR Mhealth Uhealth 2020;8(9):e18694 URL: <https://mhealth.jmir.org/2020/9/e18694> DOI: 10.2196/18694

Replication Data for: Systematic Review of the Reliability and Validity of Commercially Available Wearable Devices for Measuring Steps, Energy Expenditure, and Heart Rate. Harvard Dataverse.   URL: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/O7GQIM> [accessed 2021-03-12]

Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Yoo JY, et al. Accuracy of Fitbit Devices: Systematic Review and Narrative Syntheses of Quantitative Data. JMIR Mhealth Uhealth 2018 Aug 09;6(8):e10527 [[FREE Full text](https://mhealth.jmir.org/2018/8/e10527/)] [[CrossRef](https://dx.doi.org/10.2196/10527" \t "_blank)] [[Medline](https://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=30093371&dopt=Abstract)]