Periodicals of Engineering and Natural Sciences ISSN 2303-4521
Vol. 9, No. 3, July 2021, pp.82-89

Reliability and validity analysis of smartwatches use for healthcare

Arisa Olivia Putri [1], Musab A. M. Ali [2], Ali Abd Almisreb[3]
[1] Faculty of Engineering and Computer Science, Jakarta Global University, Jakarta, Indonesia
[2]Faculty of Engineering, Halic University, Istanbul, Turkey
[3]Faculty of Engineering and Natural Sciences, International University of Sarajevo, Sarajevo, Bosnia, and Herzegovina

ABSTRACT

The utilization of Smartwatch as a wearable sensor is one of the emergings of E-health development. It is easier to carry anywhere becomes one of the supporting factors. However, it is essential to examine the reliability and validity of smartwatches in measuring Heart Rate and Blood Pressure associated with medical uses. Methods: Eighty-eight healthy participants are recruited to be measured heart rate and Blood Pressure. The reliability and validity were determined by comparing the smartwatches with the home standard Blood Pressure using mean differences, Bland Altman plot, Interclass correlation coefficient (ICC), and Cronbach's alpha. Results: the reliability varied based on the ICC, ranging from 0.533 to 0.852. Two smartwatches showed relatively weak ICC and broad limits of agreement of the Bland–Altman plots at both heart rate and Blood Pressure Measurement. F1 Smartband Bracelet Watch showed slightly better results for heart rate measurement than Y2 Plus Smart Wrist Band.

Conversely, Y2 Plus Smart Wrist Band demonstrated the best accuracy at Systolic measurement. And for Diastolic Blood Pressure was relatively the same in reliability and validity. Conclusion: The reliability and validity of smartwatches use, especially for healthcare, are still less accurate for the clinical standard because the ICC was lower than <0.9. but for everyday use, it is good reliability and validity. For future work, the accuracy of sensor reading of smartwatches needs to improve to reach the clinical standard.

Keywords: Wearable Sensor, IoT, Validity, Reliability, Performance

Corresponding Author:

Arisa Olivia Putri
Faculty of Engineering and Computer Science
Jakarta Global University
Jakarta, Indonesia
E-mail: arisa@jgu.ac.id

1. Introduction

Nowadays, People feel more comfortable fulfilling their needs, finish their job, and do all activities using the internet. In addition, to fulfill daily needs, the development of the internet is also reaching out to the modern health sector; one of these is the development of E-health [1]. E-health helps people check their health and consult with their doctors remotely by integrating mobile technology to solve health problems (such as obesity and heart disease) and improve consumers' quality of life. One part of E-health is sensors to detect body conditions. In 2018, 431 wearable devices place on the wrist, arm, and chest developed based on the wearable database[2]. Wearable technology can measure some health parameters, such as Blood Pressure, Oxygen Level in Blood (SPO2), Heart Rate, Temperature, And Respiratory Rate. Smartwatch As wearable technology which is included in the E-health system to improve the quality of health services. This development makes it possible to provide people with information of sensing early symptoms of the disease or precaution of a diseased and keep track of the patient's condition [3]. This technology is used to detect the condition of the human body; it needs to be validated to minimize measurement errors. Analysis of reliability and validity by comparing several smartwatches that support E-health technology needs to be done to determine which Smartwatch is the most accurate to use.

2. Methods

2.1. Research method

Quantitative research is appropriate for describing phenomena by collecting numerical data analyzed utilizing mathematically based approaches to conclude [4]. This method is suitable for answering relationships within measurable variables to explain, predict and control phenomena [5]. Because this research aimed to calculate and analyze the validity and reliability of the smartwatch's measurement results, a quantitative approach was the most suitable to analyze data. The prior research about validity and reliability using the quantitative approach to produce a complete analysis data consist of a calculation table and graph [6].

2.2. Participant

The number of samples drawn is 88 samples (18-65 years) from students, men, and women who studied medicine at Management and Science University. All participants had to choose based on inclusion and exclusion criteria to key out the information relevant in answering the research question [7]. The eligible participant will be informed about the test procedure and sign a consent form before doing data collection based on the inclusion and exclusion criteria.

2.3. Instrumentation

Several instruments were used in this research: Y2 Plus Smart Wrist Band, F1 Smart band Bracelet Watch, and Omron HEM-7322 used as Home Standard Device. Y2 Plus Smart Wrist Band will be labeled as Smartwatch A, and F1 Smart band Bracelet Watch will be marked as Smartwatch B.
This study used an experimental method to obtain participant's data on heart rate and blood pressure. This research will be held in a testing laboratory, which means that participants will be observed by a medical student who joins as a committee for the MIG-WHEALTH Conference. The data using smartwatches will be collected with Omron Digital Blood Pressure in both blood pressure. Before collecting data on heart rate and Blood Pressure, the participant was asked about demographic information, for example, age, gender, height, weight, and sign the consent form. Height and Weight information got it from direct measurement before doing blood pressure measurement. Data collection will do simultaneously between the Omron Blood Pressure, Y2 Plus Smart Wrist Band, and F1 Smart band Bracelet Watch.

2.4. Statistical analysis

The data collection was collected completely to do statistical analysis. IBM SPSS Statistics 24 and Microsoft Excel are used as software for doing statistical analysis. After removing ineligible data, the filtered data will be imputed into IBM SPSS Statistic 24 software. Descriptive analysis corresponded to identifying the characteristics of an observed phenomenon or exploring correlations between two or more entities. The variable such as height, weight, gender, BMI, mean heart rate, mean systolic Blood Pressure, and diastolic Blood Pressure will be assessed.
This research will use the two types of inferential statistics: agreement and correlation, to assess validity and reliability. There are several correlation tests based on three distinct categories, nominal, ordinal, and interval. If the normality test showed the normal distribution, the parametric test could be used as correlation analysis. Pearson correlation will assess to obtain the correlation level and the significance measurement. P-value against a prespecified level of significance, which is often chosen to be 0.05. If P-value <0.05, the measurement cannot accept as highly significant, or vice versa. The Scatter Plot will assess the relationship between two variables to make it easy to understand whether the data was a positive or negative correlation. Through the calculation result and correlation graph, it can be used to measure the validity of smartwatches.
Intraclass correlation coefficient (ICC) is a widely used reliability index; in other words, it reflects the degree of correlation and the agreement between measurements. With values closer to 1 representing more substantial reliability. A two-way random-effect model based on single ratings and a whole deal will be assessed. Default 95% Confidence Intervals (CI) will be used for each ICC based on the prior studies [8]. The Bland and Altman (BA) Plot were used to assess the level of agreement between smartwatches and Omron Blood Pressure. Calculation of mean difference and limit of agreement (upper and lower) will be evaluated, and place in the

scatter plot. The reliability of the two smartwatches will be analyzed using Cronbach Alpha Coefficient [9]. The Cronbach Alpha coefficient and ICC rating scale was used to determine the reliability, as shown in Table 1.

Table 1. Rating Scale Instrument Quality Criteria [10]

Item Measurement Reliability and Correlation	
above 0.94	Excellent
0.91-0.94	Very good
0.81-0.90	Good
0.67-0.80	Fair
<0.67	Poor

3. Results

3.1. Demographic characteristic

The results showed 37.46% (33) are Male participants and 62.54% (55) are Female participants. All participants are medical students of Management and Science University that join the conference event. Most participants were normal weight, about 58.31%, 30.14% for overweight participants, and 11.56% for underweight participants. This research classified the aged range into three categories, <18 years, 19-25 years, and >25 years. The mean and standard deviation of age distribution was 24.66 years with standard deviation (SD=3.62). 68.66% of participants are aged between 19 and 25 years, and 31.34% for more than 25 years. The mean and standard deviation of body mass index was 23 kg/m2 (SD = 4.049). The mean result of weight and height, respectively were 58.41 kg (SD = 11.099), 158.52 cm (SD = 7.925)).
The mean Heart Rate for all eligible data was 87 Bpm (SD=9.18) for Omron Blood Pressure, 81 Bpm (SD=8.18) for mean Heart Rate used Smartwatch A and 87 Bpm (SD=8.77) for mean Heart rate assessed using Smartwatch B. The mean Systolic Blood Pressure was 116.53 mmHg (SD=12.575) for Omron Blood Pressure, 118.02 mmHg (SD=9.753) for Blood Pressure using Smartwatch A and 120.30 mmHg (SD=11.842) for Blood Pressure assessed using Smartwatch B. The mean Diastolic Blood Pressure was 77.48 mmHg (SD=7.662), 79.03 mmHg (SD=6.675), and 79.36 mmHg (SD=6.952) for Omron Digital Blood Pressure, Smartwatch A, and Smartwatch B.

3.2. Preliminary statistical analyses

Normality test was assessed before doing another statistical analysis. This test is essential to determine the following statistical analysis to obtain excellent and valid data. Based on the trial using Shapiro Wilk Test, the result shown that data are normally distributed.

3.3. Pearson correlation test

A correlation test was used to know the relationship between two variable scores. The Pearson Correlation Coefficient result in Heart Rate compare with Omron BP for Smartwatch A r=0.472; P<0.001 and r=0.593; P<0.001 for smartwatch B. Then, from the Pearson correlation result, smartwatch A/B and Omron BP show a significant result as P-value 0.001<0.05, which means a significant correlation between the two devices. Pearson Correlation Coefficient result in Systolic Blood Pressure A was r=0.763; P<0.001, and smartwatch B was r=0.654; P<0.001. P-value calculation of smartwatch A and smartwatch B with Omron BP show a significant result as P-value 0.001<0.05. The Pearson Correlation Coefficient result between smartwatches and Omron BP was r=0.595; P<0.001 for smartwatch A and r=0.551; P<0.001 for smartwatch B.

3.4. Intraclass correlation coefficient (ICC)

The intraclass correlation coefficient (ICC) is a widely used reliability index in reliability analyses[11]. This research used ICC(2,1) with a two-way random effect, with the absolute agreement and single

rater/measurement. The average ICC of smartwatch A is 0.564 with 95% confidence interval from 0.222 to 0.743 (F(87)= 2.769, p<.001. The average ICC of Smartwatch B compare with Omron Blood Pressure was 0.745 with a 95% confidence interval from 0.611to 0.833 (F(87)= 3.89, p<.001). The average ICC of Smartwatch A compare with Omron Blood Pressure in systolic measurement was 0.852 with a 95% confidence interval from 0.774 to 0.903 (F(87)= 6.77, p<.001). The average ICC of Smartwatch B compare with Omron Blood Pressure in systolic measurement was 0.771 with a 95% confidence interval from 0.631 to 0.855 (F(87)= 4.757, p<.001). The average ICC of Smartwatch A compare with Omron Blood Pressure in diastolic measurement was 0.733 with a 95% confidence interval from 0.592 to 0.826 (F(87)= 3.875, p<.001).

3.5. Level of agreement

Correlation quantifies the degree to which two variables are related. A high correlation does not automatically imply a good agreement between the two methods [11]. The result shows that the mean difference reaches 5.53 Bpm with limits of agreement in heart rate measurement: lower limit = −12.10 Bpm to upper limit= 23.17 Bpm. Then, the mean difference was -1.01 mmHg and limits of agreement −16.63 mmHg to 14.61 mmHg for Systolic Blood Pressure. The Bland Altman for Diastolic shows the same result; the mean difference was -1.54; the limits of agreement were again too wide (−14.30 mmHg to 11.21 mmHg) for the level of agreement based on comparison Blood Pressure range of within 10 mmHg and 20 mmHg for Diastolic.

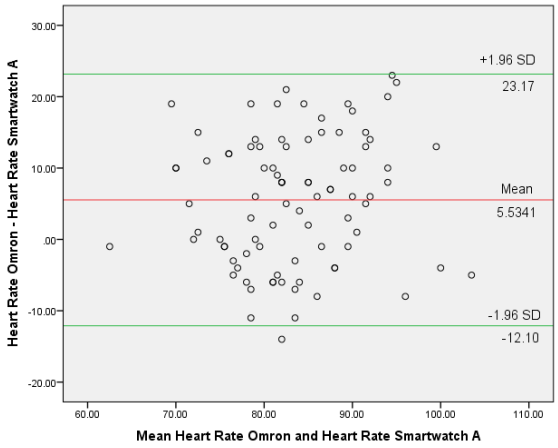

Figure 1. Bland Altman plot for Heart Rate Blood Pressure Omron Blood Pressure and Smartwatch A

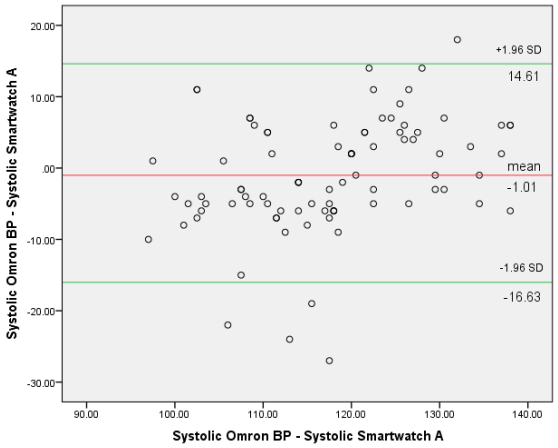

Figure 2. Bland Altman plot for Systolic Blood Pressure Omron Blood Pressure and Smartwatch A

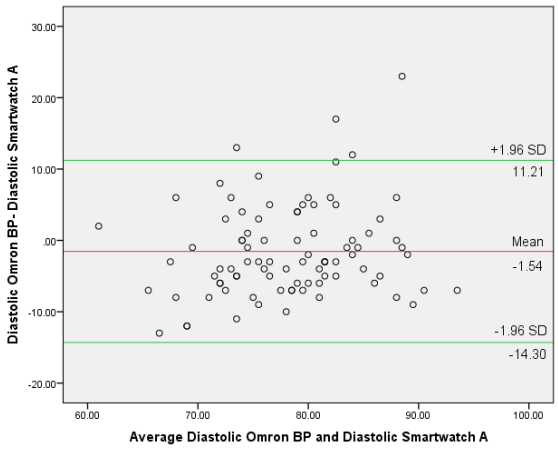

Figure 3. Bland Altman plot for Diastolic Blood Pressure Omron Blood Pressure and Smartwatch A

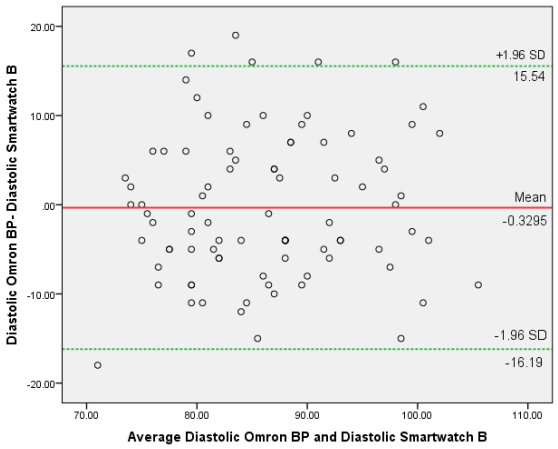

Figure 4. Bland Altman plot for Heart Rate Omron Blood Pressure and Smartwatch B

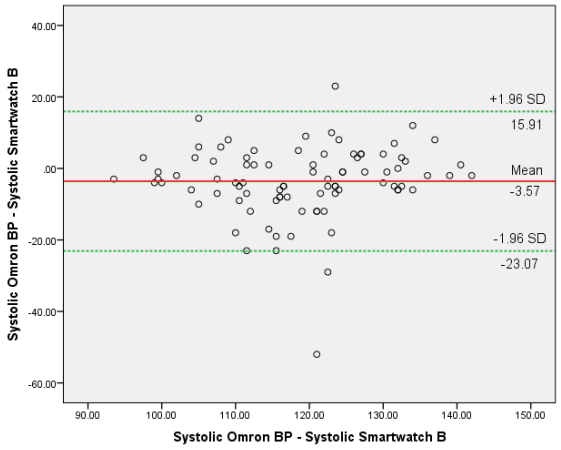

Figure 5. Bland Altman plot for Systolic Blood Pressure Omron Blood Pressure and Smartwatch B

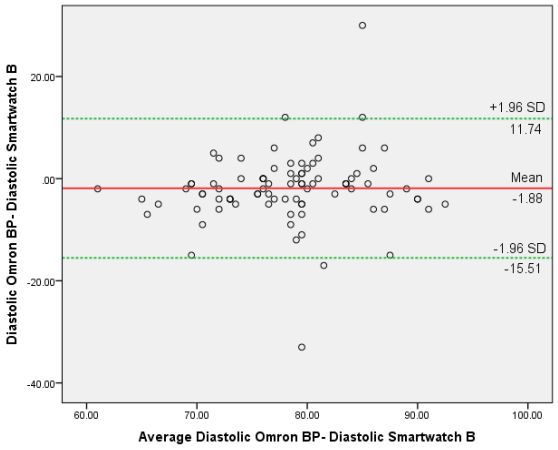

Figure 6. Bland Altman plot for Diastolic Blood Pressure Omron Blood Pressure and Smartwatch B

The result shows that the mean difference reaches -0.32 Bpm with limits of agreement in heart rate measurement: upper limit = 15.54.10 Bpm to lower limit= -16.19 Bpm. Then, the mean difference was -3.57 mmHg and limits of agreement −23.07 to 15.91 mmHg for Systolic Blood Pressure. The Bland Altman for Diastolic shows the same result; the mean difference was -1.88 the limits of agreement were again too wide (−15.51 to 11.74 mmHg) for the level of agreement based on comparison Blood Pressure range of within 10 mmHg and 20 mmHg for Diastolic [12]. The limit of the agreement was too wide for each variable, heart rate, systolic and diastolic.

3.6. Reliability analysis.

One of the critical issues of reliability concerns the scale's consistency. One of the most commonly used indicators of internal consistency is Cronbach's alpha coefficient (α). To obtain Cronbach's alpha, the calculation using formula (1) was used. The comparison result of Reliability between Smartwatch A and Smartwatch B is shown in Table 2.

Table 2. Reliability between smartwatch A and B with Omron Blood Pressure, using Cronbach Alpha

Blood Pressure	Method of Comparison	Cronbach Alpha
Systolic Blood Pressure	Smartwatch A and Omron BP	0.852
	Smartwatch B and Omron BP	0.790
Diastolic Blood Pressure	Smartwatch A and Omron BP	0.742
	Smartwatch B and Omron BP	0.708
Heart Rate	Smartwatch A and Omron BP	0.639
	Smartwatch B and Omron BP	0.744

$$\alpha = \frac{k\ X\ \bar{c}}{\bar{v} + (k-1)\bar{c}} \qquad (1)$$

Where α= Cronbach alpha

k = the number of scale items

$\bar{c}$ =the average of all covariances between items

$\bar{v}$ = the average variance of each item

Using equation (1), the calculation of Cronbach's alpha has done for all variables, heart rate, systolic, and diastolic blood pressure. The calculation result was shown in Table 2; the Smartwatch A shows the Cronbach Alpha reliability reach α =0.852 for Systolic Blood Pressure and α =0.742 for Diastolic Blood Pressure. Heart rate α have a weak to fair reliability with α = 0.639. The smartwatch B showed the Cronbach Alpha reliability reach 0.790 for Systolic Blood Pressure and 0.708 for Diastolic Blood Pressure.

4. Discussion

4.1. Reliability of smartwatch measurement

The reliability of the smartwatch measurement was assessed using the Cronbach Alpha coefficient. As described in Table 2, systolic Blood Pressure has the highest degree of reliability(α=0.852), and heart rate measurement has the lowest degree of reliability (=α 0.639). The reliability of diastolic measurement in both smartwatch A and smartwatch B were relatively same. For statistical analysis, it is quite good if the reliability more than 0.6 [9]. It could be concluded based on statistical analysis of reliability using Cronbach Alpha. The results show pretty satisfactory results; however, for healthcare, the level of reliability still needs to be increased. Nevertheless, the current work still less than the clinical standard. The alpha number should at least be 0.90, and then the Smartwatch can be stated as a valid device for healthcare use.

4.2. Validity of smartwatch measurement

There are several ways to assess the validity of smartwatches in laboratory conditions. Based on [6], there are three methods for determining the validity, assess systematic difference, examine Interclass Correlation Coefficient (ICC), and assess the level of agreement. This research used two of three methods to investigate the Interclass Correlation Coefficient (ICC) and the level of agreement to assess the level of validity.

The use of correlation coefficients is the most common method of analyzing this type of medical result. Pearson correlation coefficient result of smartwatch A and Omron Blood Pressure showed a weak correlation in heart rate measurement. The Pearson correlation coefficient result for heart rate is still lower than 0.6.

The Pearson correlation was r=472; P<0.001 and smartwatch B r=593; P<0.001. Diastolic Blood Pressure measurement also has a weak Pearson correlation in two smartwatches measurements, but smartwatch A was slightly better than smartwatch B. For the Systolic Blood Pressure, smartwatch A tends to have a reasonable correlation, assessed using Pearson correlation coefficient or Interclass correlation coefficient. The result was r=0.763, P < 0.001 for comparison of systolic blood pressure Smartwatch A and Omron Blood Pressure and r=0.654, P<0.001 for comparison systolic blood pressure, smartwatch B and Omron Blood Pressure. The result was r=0.595, P < 0.001 for comparison of systolic blood pressure Smartwatch A and Omron Blood Pressure and r=0.551, P<0.001 for comparison systolic blood pressure, smartwatch B and Omron Blood Pressure.

ICC result of smartwatch A and B showed that smartwatch B was better than A, resulting from ICC smartwatch B 0.745 with P<0.001 in heart rate measurement. Despite the lower ICC of smartwatch A, the P-value of Smartwatch A is also significant (P<0.001).ICC result of smartwatch A and B showed that smartwatch A was better than B, Smartwatch A ICC=0.852 with P<0.001 in systolic measurement and diastolic measurement ICC=0.733 with P<0.001.

The other analysis was used to convince the validity degree. The level of agreement of Bland Altman plots is used to assess the difference between lower and upper limits of agreement [6]. Some research has given the wide limits of agreement that neither is sufficient to predict ambulatory Blood Pressure [13]. The significant finding was considerable variability between the two types of Blood Pressure measurements, with wide limits of agreement for Systolic and Diastolic Blood Pressure.

It shows in Figure 2, the mean difference for Smartwatch A compares with Omron Blood Pressure was -1.01 mmHg, but the limits of agreement are still wide, −16.63 mmHg to 14.61 mmHg for Systolic Blood Pressure. The same result is shown in Figure 5. The agreement between Omron Blood Pressure and Smartwatch B, the mean difference was -3.57 mmHg higher than Smartwatch A, with limits of agreement of −23.07 mmHg to 15.91 mmHg for Systolic Blood Pressure. The level of agreement was assessed between smartwatches, and Omron Blood Pressure showed a wide range in heart rate; both smartwatch A and smartwatch B calculated more

than 30 Blood Pressure for individual measurement. It is the same with the analysis of the previous research in heart rate measurement using Fitbit charge 2 [14]. Blood Pressure measurement also gets a wide limit of agreement, as the standard set 10 mmHg difference for diastolic and 20 mmHg for Systolic Blood Pressure [15]. The validity heart rate measurement of the two smartwatches is weak, but the validity is good for Blood Pressure. However, the clinical validity cannot accept the measurement result. The statistical analysis showed below the clinical standard indicated in the Bland Altman Plot, as shown in Figure 1-6. As stated by [14], one aspect that makes the wide range limit of agreement was indicated the un proper placement of device would affect the measurement result.

5. Conclusion and future work

A prominent accuracy level is needed for E-health usage. This research shows that Pearson's correlations between Smartwatch A or Smartwatch B with Omron Blood Pressure measurements were high. Still, it is inadequate when measuring heart rate, the Pearson Correlation Coefficient can reach. In the same way, the reliability of Smartwatch Assesses using Cronbach's Alpha coefficient, and the result shows good reliability as showed that calculation was between 0.6-0.8. the best Smartwatch between two devices for heart rate measurement was smartwatch B, and even the statistical result showed a slightly better number. And for blood pressure validity and reliability, this research concludes that the best was Smartwatch A. But, for clinical purposes or healthcare, two smartwatches were still less accurate as the standard for validity and reliability were must be more than 0.9. In future work, examine the Blood Pressure and Heart Rate of other wearable devices in different physical activities, comparing to the Gold Standard Devices in Hospital, is needed. Then, we can know physical activities can impact the reliability and validity of wearable devices for E-Health.

5.1. Ethical considerations

Ethic was submitted to the Research Management Center (RMC) of Management and Science University to conduct the research. Permission was obtained from the Committee of Migrant Worker Health Research Network (MIG-WHEALT) Conference to join the event and to collect data for this research. Informed written consent was obtained from the participants of the study. The participation was completely voluntary. Subjects were given the freedom to withdraw from the study at any time. Confidentiality of the collected data was maintained.

References

[1] S. S.-L. Tan and N. Goonawardene, "Internet Health Information Seeking and the Patient-Physician Relationship: A Systematic Review.," *J. Med. Internet Res.*, vol. 19, no. 1, p. e9, Jan. 2017, doi: 10.2196/jmir.5729.

[2] Vandrico Solutions Inc., "Wearable Technology Database | Vandrico Inc.,." https://vandrico.com/wearables.html (accessed Aug. 27, 2018).

[3] G. Eysenbach, "What is e-health?," *J. Med. Internet Res.*, vol. 3, no. 2, p. e20, Jun. 2017, doi: 10.2196/jmir.3.2.e20.

[4] D. Muijs, *Doing quantitive research in education with spss by daniel muijs*. 2004.

[5] I. Rajagopal and N. Bojin, "Quantitative Reserach Methods," *First Monday*, vol. 8, no. 1, 2003, doi: http://dx.doi.org/10.5210/fm.v8i1.1023.

[6] T. J. M. Kooiman, M. L. Dontje, S. R. Sprenger, W. P. Krijnen, and C. P. Van Der Schans, "Reliability and validity of ten consumer activity trackers," *BMC Sports Sci. Med. Rehabil.*, pp. 1–11, 2015, doi: 10.1186/s13102-015-0018-5.

[7] C. M. Patino and J. C. Ferreira, "Inclusion and exclusion criteria in research studies: definitions and why they matter," *J. Bras. Pneumol.*, vol. 44, no. 2, pp. 84–84, 2018, doi: 10.1590/s1806-37562018000000088.

[8] T. A. Floegel, A. Florez-Pregonero, E. B. Hekler, and M. P. Buman, "Validation of Consumer-Based Hip and Wrist Activity Monitors in Older Adults With Varied Ambulatory Abilities," *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.*, vol. 72, no. 2, pp. 229–236, Feb. 2017, doi: 10.1093/gerona/glw098.

[9] J. M. Bland and D. G. Altman, “Statistics notes: Cronbach’s alpha,” *BMJ*, 1997, doi: 10.1136/bmj.314.7080.572.

[10] M. M. Mohamad, N. L. Sulaiman, L. C. Sern, and K. M. Salleh, "Measuring the Validity and Reliability of Research Instruments," *Procedia - Soc. Behav. Sci.*, vol. 204, no. November 2014, pp. 164–171, 2015, doi: 10.1016/j.sbspro.2015.08.129.

[11] T. K. Koo and M. Y. Li, "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research," *J. Chiropr. Med.*, vol. 15, no. 2, pp. 155–163, 2016, doi: 10.1016/j.jcm.2016.02.012.

[12] C. E. Clark, R. S. Taylor, A. C. Shore, and J. L. Campbell, "The difference in blood pressure readings between arms and survival: Primary care cohort study," *BMJ*, vol. 344, no. 7851, pp. 1–13, 2012, doi: 10.1136/bmj.e1327.

[13] D. I. D. Ona, R. G. Sy, D. D. Morales, A. V Aquino, and S. Nelson, "Validation of the Omron automated oscillometric blood pressure monitoring devices (HEM 7130 / HEM 7120 / HEM 7322)," pp. 1–8, 2008.

[14] S. Benedetto, C. Caldato, E. Bazzan, D. C. Greenwood, V. Pensabene, and P. Actis, "Assessment of the fitbit charge 2 for monitoring heart rate," *PLoS One*, vol. 13, no. 2, pp. 1–10, 2018, doi: 10.1371/journal.pone.0192691.

[15] G. Mlawanda, M. Pather, and S. Govender, "An analysis of blood pressure measurement in a primary care hospital in Swaziland," *African J. Prim. Heal. Care Fam. Med.*, vol. 6, no. 1, pp. 1–9, 2014, doi: 10.4102/phcfm.v6i1.590.