Journal Pre-proofs

Benchmarking between two wearable inertial systems for gait analysis based on a different sensor placement using several statistical approaches

Leandro Donisi, Gaetano Pagano, Giuseppe Cesarelli, Armando Coccia, Federica Amitrano, Giovanni D'Addio

PII: S0263-2241(20)31158-1
DOI: https://doi.org/10.1016/j.measurement.2020.108642
Reference: MEASUR 108642

To appear in: *Measurement*

Received Date: 9 May 2020
Revised Date: 31 August 2020
Accepted Date: 18 October 2020

Please cite this article as: L. Donisi, G. Pagano, G. Cesarelli, A. Coccia, F. Amitrano, G. D'Addio, Benchmarking between two wearable inertial systems for gait analysis based on a different sensor placement using several statistical approaches, *Measurement* (2020), doi: https://doi.org/10.1016/j.measurement.2020.108642

Benchmarking between two wearable inertial systems for gait analysis based on a different sensor placement using several statistical approaches

Leandro Donisi[1,2], Gaetano Pagano[3], Giuseppe Cesarelli[4,5], Armando Coccia[2,6], Federica Amitrano[2,6], Giovanni D'Addio[2]

[1]Department of Advanced Biomedical Sciences, University of Naples Federico II, Naples, Italy
[2]Bioengineering Unit, ICS Maugeri SPA SB Institute of Care and Scientific Research, Telese Terme (BN), Italy
[3]Bioengineering Unit, ICS Maugeri SPA SB Institute of Care and Scientific Research, Bari, Italy
[4]Department of Chemical, Materials and Production Engineering, University of Naples Federico II, Naples, Italy
[5]Center for Advanced Biomaterials for Healthcare, Italian Institute of Technology, Naples, Italy
[6]Department of Electrical Engineering and Information Technology, University of Naples Federico II, Naples, Italy

Corresponding Author:
Giovanni D'Addio
E-mail address: gianni.daddio@icsmaugeri.it
Full Postal Address: Via Bagni Vecchi, 1, 82037, Telese Terme (BN), Italy

Abstract

Despite the growing use of different wearable inertial systems for gait analysis in clinical setting, also based on a different sensor placement, there is still a lack of knowledge about the agreement between them and their repeatability. The purpose of this study is to investigate the agreement between two commercial wearable inertial systems for gait analysis: Opal and G-Walk Systems, and their repeatability. Fifty-three subjects, healthy and pathological, underwent a gait analysis session instrumented by both systems, seven spatiotemporal parameters were recorded. The study of agreement was carried out through Bland-Altman Analysis, Passing-Bablok regression and Paired t-test, the study of repeatability through the intra class correlation coefficient ICC(3,1). Study results showed a not perfect agreement between the two systems although they both showed good repeatability. This work underlines the importance to perform a study of agreement before using devices interchangeably or even as a replacement in order to have reliable measurements.

Keywords: wearable devices, IMUs, gait analysis, biomechanics, agreement, repeatability.

1 Introduction

The evaluation of rehabilitation outcome in terms of spatiotemporal and kinematic parameters is of primary importance in the clinical field to assess the degree of functional and motor recovery following pharmacological interventions and physiotherapy. Several studies have focused on motor function assessments [1,2] and rehabilitation medicine [3]. Measurement of temporo-spatial gait parameters are used by physicians and physiotherapists for diagnosing gait disorders, disease progression monitoring and controlling the effect of different therapeutic methods [4]. Individuals with gait disorders in fact show an ample range of spatiotemporal and kinematic variations, that lead the values to move away from normal ranges. Understanding these abnormalities in the human movement is critical for therapeutic planning, management and clinical decision making. In the clinical setting, therapists often evaluate gait, mobility and balance using visual observations [5], questionnaires or scales [6,7] and functional assessment interviews [8] to establish abnormalities in kinematics. While these methods of gait and mobility assessments are simple and do not require technologically advanced instruments, they are often qualitative and are incline to evaluator bias [9]. Further, they are operator-dependent and often miss inter-rater reliability [10-13]. Alternative, objective assessments are critical for accurate documentation of patients progress and are increasingly sought by third-party payers [9]. Modern biomedical technologies, in fact, are increasingly enabling to support the use of qualitative methods deriving from clinical scales. There are several measurement systems used in the gait analysis field to assess gait and balance quantitatively. Undoubtedly the gold standard is represented by three-dimensional motion capture systems and force plates but these tools are onerous to acquire and operate, reducing their feasibility for clinical purpose. There are several low-cost instruments such as: e-textile socks [14], Kinect [15], Wii fit [16] and even webcams [17], which are appealing for clinicians to perform quantitative assessments despite their intrinsic limitations. Their limits are that all of them are restricted to a small capture volume, and only few steps are available for data analysis procedures. Recently the development and the spread of inertial measurement units (IMUs) for spatiotemporal and kinematic assessments have been an important and innovative technological breakthrough in the field of biomechanics and wearable sensors

[18, 19]. The reason for their success and diffusion is due to multiple factors as they are relatively inexpensive, allow a virtually unlimited number of steps to be evaluated, present a lower complexity of the experimental setup and of the data processing procedures, need a limited time for the examination, finally provide the ability to evaluate gait and balance disorders outside the limited space of the clinical and research laboratory [20]. The two wearable inertial systems compared in the present study are based on IMUs, and their advantages showed above have led them to have a very increasingly use in clinical practice and research; different studies present in the scientific literature in fact are based on the use of wearable inertial systems to monitor gait in different occurrences [21]. Despite the growing use of different wearable inertial systems for gait analysis in the clinical setting, there is still a lack of knowledge about the agreement between systems that although they use the same technology, in this case IMUs, are based on a different sensor placement. Basing on literature available results in fact neither a standard protocol for validation nor fixed metrological characteristics can be identified, wearable devices are validated without standard procedures (test protocol, population characteristics and metrological parameters) which turns into irregular results, barely comparable each other [22]. The working principle of these sensors is based on the measurement of inertia and can be applied anywhere without a reference [23]. On the market in fact there are a lot of wearable inertial systems for gait analysis based on a different sensor placement as the two systems analyzed in this study. The primary purpose of this work was to study the agreement between two commercial wearable inertial systems: Opal System by APDM Inc. and G-Walk System by BTS Bioengineering Inc. widely used in the clinical and rehabilitation medicine fields to assess the rehabilitation outcomes of patients with motor pathologies and in the research field [24,25]. Different spatiotemporal parameters related to the gait were considered for the analysis. Moreover, a repeatability study for each measurement system and for each motion parameter was performed.

2 Materials and Methods

2.1 Wearable Inertial Systems for Gait Analysis

Two commercial wearable inertial systems for gait analysis: Opal System by APDM Inc. and G-Walk System by BTS Bioengineering Inc. were compared in this study (Figure 1).
Opal System is composed of 6 movement monitors or opal sensors. Each sensor includes a 3 axes accelerometer with 14 bits resolution and selectable ± 16 g or ± 200 g ranges, a 3 axes gyroscope with 16 bits resolution and ± 2000 deg/s range; a 3 axes magnetometer with 12 bits resolution and ± 8 Gauss range. Opals sensors are attachable on subjects using a selection of straps and wireless connected by Bluetooth communication protocol 3.0 to a remote laptop running the Mobility Lab Software that allows to configure the hardware, record and process all movement data. Two movement monitors placed on both shins allow to calculate all the spatiotemporal parameters related to lower limb gait.
G-Walk System is based on a single wearable device namely G-Sensor equipped with 4 inertial platforms run by Sensor Fusion technology, each of them including a 16-bit triaxial accelerometer with a selectable sensitivity up to ± 16 g, a 16-bit triaxial gyroscope with a selectable sensitivity up to 2000 deg/s, a 13-bit triaxial magnetometer with a sensitivity of 1200 µT. G-Sensor is in wireless Bluetooth 3.0

communication with a PC running the G-Studio software able to store and to compute the main spatiotemporal parameters related to the gait.

Figure 1 - GWALK System on the left and OPAL System on the right

2.2 Study Population

The study was carried out on a wide population including patients with different motor pathologies and healthy subjects in order to make the sample varied and to evaluate the performances of the two devices on a wide range of values, excellent for carrying out an exhaustive benchmarking study. The study population was composed of 53 subjects. Twenty-one pathological subjects (age: 49.5 ± 17.4), of whom 13 males, with motor disabilities such as: Parkinson's disease, Stroke, Ataxia, Hemiparesis, Hemiplegia, Neuropathy, Multiple Sclerosis, Spastic Paraparesis under rehabilitation at the Institute of Care and Scientific Research ICS Maugeri SPA SB of Telese Terme (BN) in Italy. Patients were able to perform a gait analysis session without any support or interruption. Thirty-two healthy subjects (age: 45.4 ± 14.2), of which 17 males, with no signs of neurological or orthopedic impairment. All patients and all healthy volunteer signed the informed consent, the local Ethics Committee approved the study, which was performed in accordance with the Declaration of Helsinki.

2.3 Study Protocol

All subjects underwent a gait analysis session instrumented by both wearable inertial systems: 2 Opal sensors placed on each shin by velcro straps and 1 G-Sensor placed in the pocket of a belt in a lumbosacral region between S1-S2 vertebra in according with related handbooks (Figure 2). The gait analysis protocol consisted in two consecutive trials of a 10 meters walk along a straight path, spaced by a pause of at least 5 minutes. For each subject the mean value between the two trials for each spatiotemporal parameter was computed for the analysis of agreement; for the analysis of repeatability the single value of each motion parameter of both trials was considered. The following spatiotemporal parameters (STP) were analyzed: Cadence CA (steps/minute): stepping rate; Stride Velocity SV (m/s): walking speed; Gait Cycle Time GCT (s): duration of a complete gait cycle; Stride Length SL (m): distance between two consecutive foot falls at the moments of initial contacts; % Stride Length %SL (% stature): distance between two consecutive foot fall at the moment of initial contact as percentage of the subject's stature; Stance Phase STP (% gait cycle time): average percentage of a gait cycle that either foot is on the ground; Swing Phase SWP (% gait cycle time): average percentage of a gait cycle that either foot is off the ground.

Figure 2 - Sensor Placement: 2 Opal sensor attached on each shin and 1 G-Sensor placed on lumbosacral region

2.4 Statistical Analysis

Agreement between motion measurements computed by the two wearable inertial systems: Opal and G-Walk was investigated by Bland-Altman Analysis, Passing-Bablok linear regression and two-tailed paired t-test; the latter in its parametric or non-parametric form (by Wilcoxon signed-rank test) in according to D'Agostino-Pearson omnibus normality test result. Moreover, Pearson's correlation coefficient (r) was used to measure the linear strength of association between each spatiotemporal parameter computed by the two systems: Opal and G-Walk. Pearson's correlations are a poor indicator of validity since they do not account for absolute agreement, but they indicate if measurements can be fixed with recalibration (i.e., if a variable has a high Pearson's correlation, a scaling or offset can be applied to allow absolute agreement).

Bland-Altman plot describes the bias and the standard deviation respectively as the average and the standard deviation of the differences between the measures computed by two measurement methods. This method allows to calculate the limits of agreement, computed as the bias ± 1.96 times its standard deviation [26-28] and study the absence of a significant statistical error, depending if the confidence limits of the bias contain or not 0 value [29,30].

Passing-Bablok linear regression analysis allows to evaluate the constant systematic error (intercept of regression line) and the proportional systematic error (slope of regression line) according if their related 95% confidence limits include or not 0 value for the intercept and 1 for the slope [29,31].

Finally Paired t-test was used to reject the null hypothesis (there is no difference between the two systems in mean values of each spatiotemporal parameter), through the calculation of the p-value; a two-tail test was used and the nominal alpha level was set to 0.05 [32].

Repeatability or test-retest reliability of each measurement system was investigated by means of the Intraclass Correlation Coefficient (ICC). There are different forms of ICC that can give different results when applied to the same set of data. McGraw and Wong defined 10 forms of ICC based on the model (1-way random effects, 2-way random effects, or 2-way fixed effects), the type (single rater/measurement or the mean of k raters/measurements), and the definition of relationship considered to be important (consistency or absolute agreement) [33]. Shrout and Fleiss defined 6 forms of ICC, and they are presented as 2 numbers in parentheses; the first number refers to the model (1, 2 or 3) and the second number refers to the type, which is either a single rater/measurement (1) or the mean of k raters/meaurements (k) [34]. A Shrout and Fleiss convention was used in this work. Our setting let us to choice as model a two-way Mixed-Effects model and as a type 1, and so a ICC(3,1). Koo and Lii suggest to interpret ICC as poor when less than 0.5, moderate between 0.50 and 0.75, good between 0.75 and 0.90 and excellent when above 0.90 [35].

3 Results

Concerning the analysis of agreement between the two systems in relation to each spatiotemporal parameter analyzed in the study, results are shown in Tables 1, 2, 3 and Figures 3, 4, 5, 6, 7, 8, 9. Table 1 shows the results of Bland-Altman analysis where: bias is the mean of differences between the measures computed by the two systems, made for each motion parameter; LBb is the 95% lower bound bias, UBb is the 95% upper bound bias, LBb and UBb represents the limits of the 95% confidence interval of the bias; LB is the Bland-Altman lower bound and UB is the Bland-Altman upper bound, LB and UB represent the limits of agreement of the Bland-Altman Analysis. Table 2 shows the results of Passing-Bablok linear regression analysis where: m is the slope of regression line, LBm is the slope lower bound, UBm is the slope upper bound, LBm and UBm represents the limits of the 95% confidence interval of the slope; q is the intercept of regression line (offset); LBq is the intercept lower bound, UBq is the intercept upper bound, LBq and UBq represent the limits of the 95% confidence interval of the intercept. Moreover the Pearson's correlation coefficient (r) value was reported in Table 2. Table 1 and Table 2 are associated with related plots showed in the Figures 3, 4, 5, 6, 7, 8, 9. Table 3 shows the results of Paired t-test where for each spatiotemporal parameter and for each system is reported the mean value and its standard deviation and a symbol (*) in accordance with the following convention: * $p<0.5$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$ (where p is the p-value), a two tailed paired t-test with a confidence interval of 95 % was performed.

Concerning test-retest reliability or repeatability analysis for each system and for each spatiotemporal parameter, results are shown in Table 4, where the ICC(3,1) values for each parameters and for both systems are given.

Note: in the Bland-Altman plot the continuous red line represents the bias, the two dotted red line define the 95% confidence interval of the bias, the two dotted black line define the limits of agreement; in the Scatter Plot the continuous red line represents the line of identity and the blue one the Passing-Bablok regression line.

Table 1 – AGREEMENT: Bland-Atman Analysis

STP	bias	LBb	UBb	LB	UB
CA	-1.59	-2.31	-0.88	-6.80	3.61
SV	0.13	0.09	0.16	-0.12	0.37
GCT	0.00	0.00	0.01	-0.02	0.03
SL	0.14	0.11	0.18	-0.14	0.42
%SL	8.37	6.16	10.6	-7.79	24.6
STP	2.03	0.86	3.19	-6.45	10.5
SWP	-2.02	-3.19	-0.86	-10.5	6.45

Table 2 – AGREEMENT: Passing-Bablok regression Analysis

STP	r	m	LBm	UBm	q	LBq	UBq
CA	0.98	0.96	0.93	1.00	5.03	0.84	9.22
SV	0.89	1.36	1.15	1.58	-0.60	-0.90	-0.32
GCT	0.99	0.98	0.95	1.01	0.03	-0.01	0.05
SL	0.80	1.45	1.11	1.83	-0.76	-1.29	-0.28
SL%	0.80	1.58	1.25	1.99	-56.2	-90.4	-29.2
STP	-0.05	0.68	0.33	1.17	18.5	-11.7	39.8
SWP	-0.05	0.68	0.34	1.17	13.5	-5.25	26.7

Table 3 – AGREEMENT: Paired t-test Analysis

STP	OPAL	G-WALK
CA	109 ± 13	111 ± 12****
SV	1.27 ± 0.23	1.14 ± 0.27****
GCT	1.11 ± 0.16	1.11 ± 0.16
SL	1.38 ± 0.20	1.23 ± 0.24****
%SL	81.0 ± 11	72.6 ± 14****
STP	62.2 ± 3.7	60.1 ± 2.0***
SWP	37.8 ± 3.7	39.9 ± 2.0***

* p<0.5, ** p<0.01, *** p<0.001, **** p<0.0001 at T-test

Table 4 – REPEATABILITY – ICC(3,1)

STP	ICC(3,1) OPAL	ICC(3,1) G-WALK
CA	0.97	0.95
SV	0.97	0.93
GCT	0.96	0.97
SL	0.98	0.89
%SL	0.98	0.89
STP	0.96	0.82
SWP	0.96	0.82

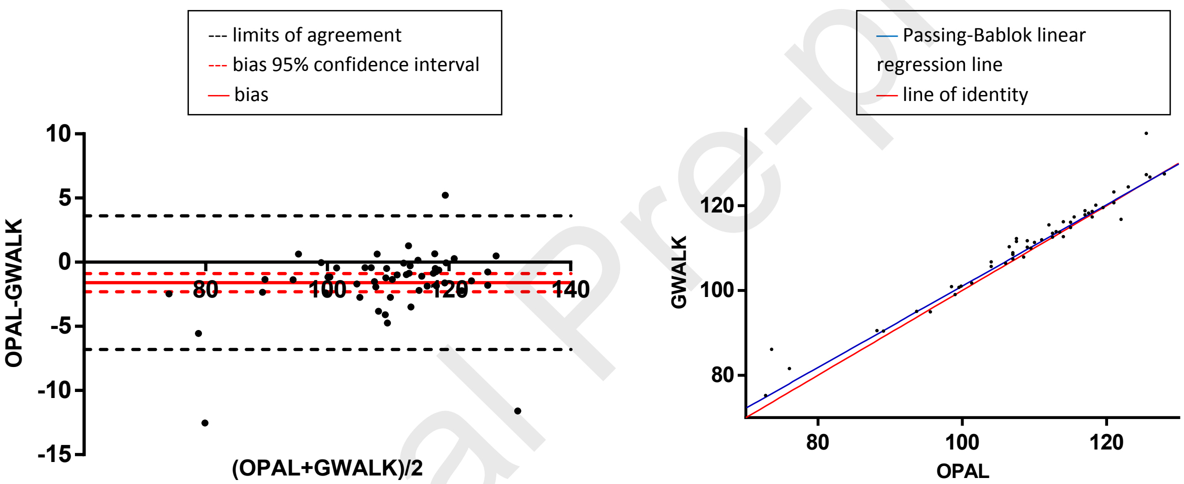

Figure 3 – Bland-Altman plot (difference vs average) of the CADENCE parameter on the left - bias computed as the mean of differences, bias 95% confidence interval computed as $bias \pm 1.96\frac{SD}{\sqrt{N}}$, limits of agreement computed as $bias \pm 1.96\,SD$ where SD is the standard deviation of the differences and N is the sample size
Scatter plot of the CADENCE parameter on the right - line of identity: GWALK = OPAL, Passing-Bablok linear regression line: linear regression line computed according the Passing-Bablok method

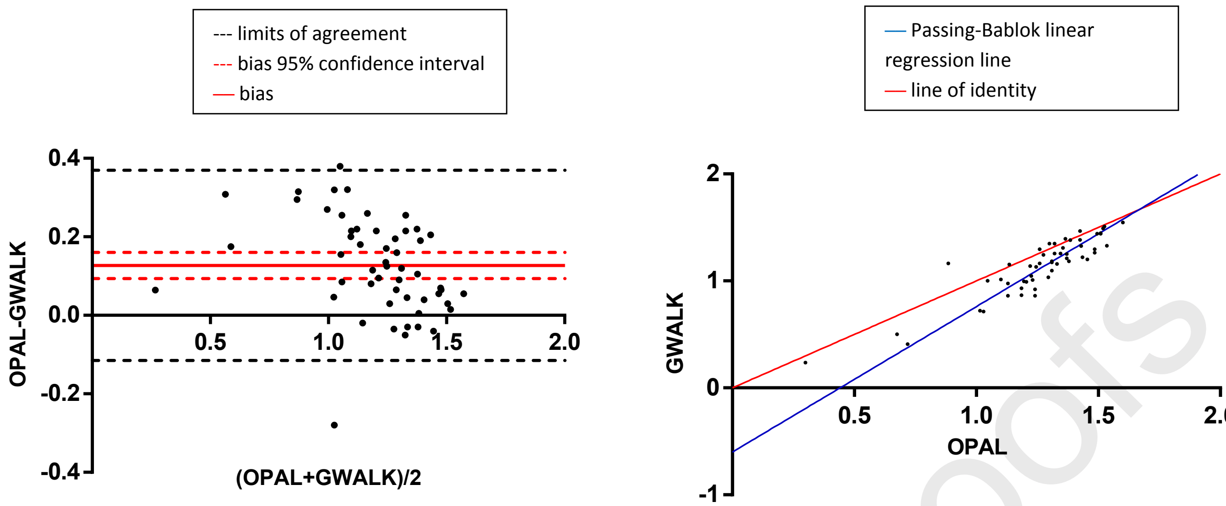

Figure 4 – Bland-Altman plot (difference vs average) of the STRIDE VELOCITY parameter on the left - bias computed as the mean of differences, bias 95% confidence interval computed as $bias \pm 1.96\frac{SD}{\sqrt{N}}$, limits of agreement computed as $bias \pm 1.96\ SD$ where SD is the standard deviation of the differences and N is the sample size
Scatter plot of the STRIDE VELOCITY parameter on the right - line of identity: GWALK = OPAL, Passing-Bablok linear regression line: linear regression line computed according the Passing-Bablok method

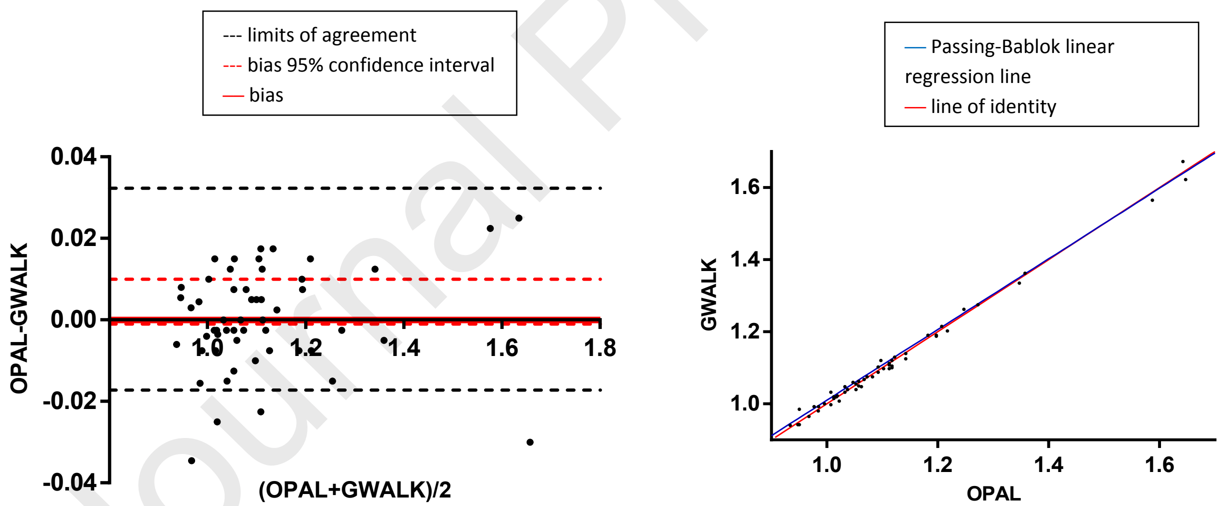

Figure 5 – Bland-Altman plot (difference vs average) of the GAIT CYCLE TIME parameter on the left - bias computed as the mean of differences, bias 95% confidence interval computed as $bias \pm 1.96\frac{SD}{\sqrt{N}}$, limits of agreement computed as $bias \pm 1.96\ SD$ where SD is the standard deviation of the differences and N is the sample size
Scatter plot of the GAIT CYCLE TIME parameter on the right - line of identity: GWALK = OPAL, Passing-Bablok linear regression line: linear regression line computed according the Passing-Bablok method

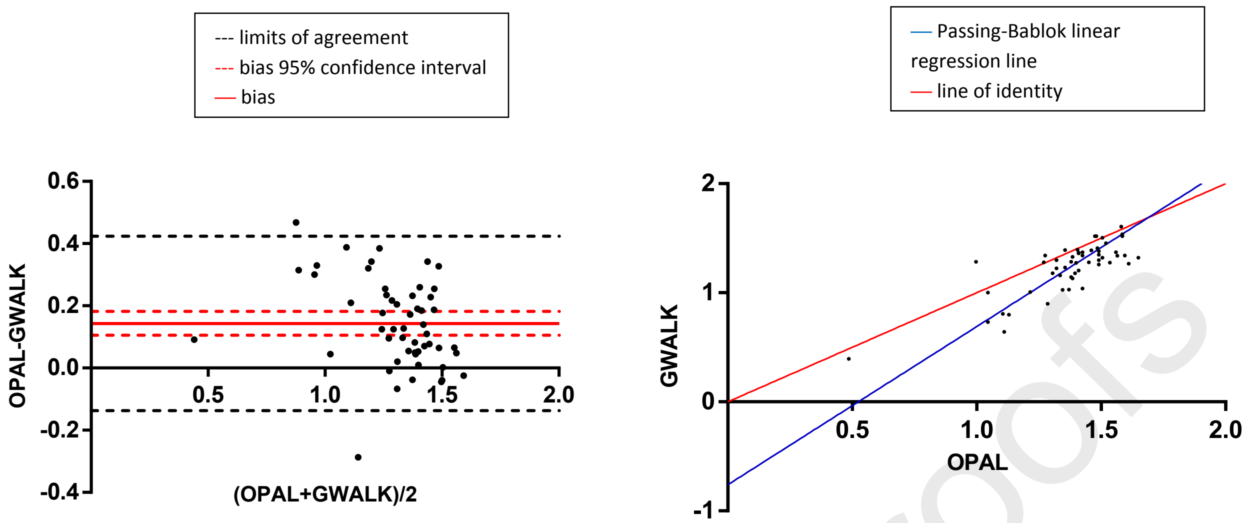

Figure 6 – Bland-Altman plot (difference vs average) of the STRIDE LENGTH parameter on the left - bias computed as the mean of differences, bias 95% confidence interval computed as $bias \pm 1.96\frac{SD}{\sqrt{N}}$, limits of agreement computed as $bias \pm 1.96\, SD$ where SD is the standard deviation of the differences and N is the sample size

Scatter plot of the STRIDE LENGTH parameter on the right - line of identity: GWALK = OPAL, Passing-Bablok linear regression line: linear regression line computed according the Passing-Bablok method

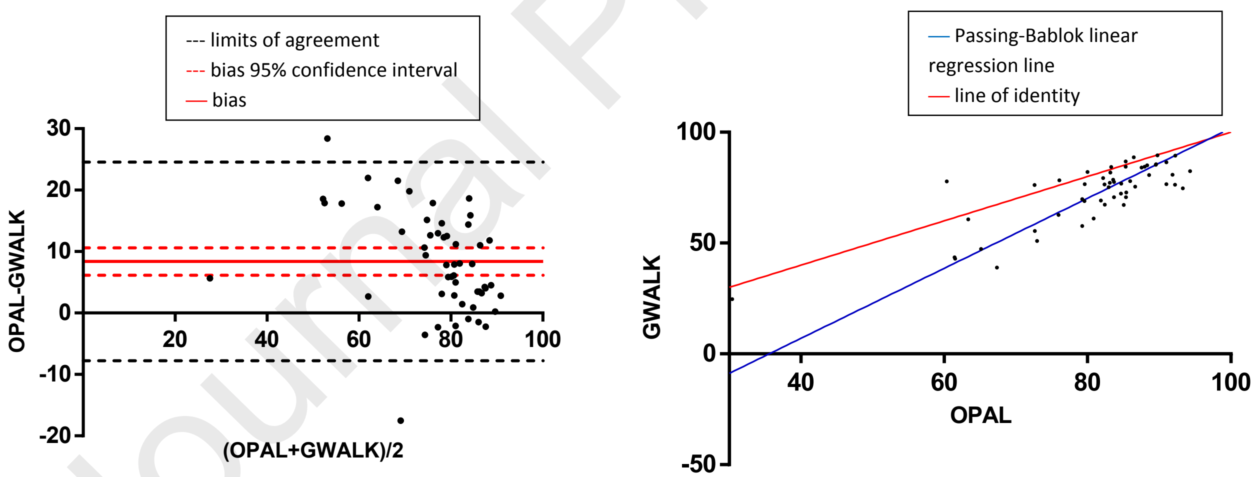

Figure 7 – Bland-Altman plot (difference vs average) of the STRIDE LENGTH % parameter on the left - bias computed as the mean of differences, bias 95% confidence interval computed as $bias \pm 1.96\frac{SD}{\sqrt{N}}$, limits of agreement computed as $bias \pm 1.96\, SD$ where SD is the standard deviation of the differences and N is the sample size

Scatter plot of the STRIDE LENGTH % parameter on the right - line of identity: GWALK = OPAL, Passing-Bablok linear regression line: linear regression line computed according the Passing-Bablok method

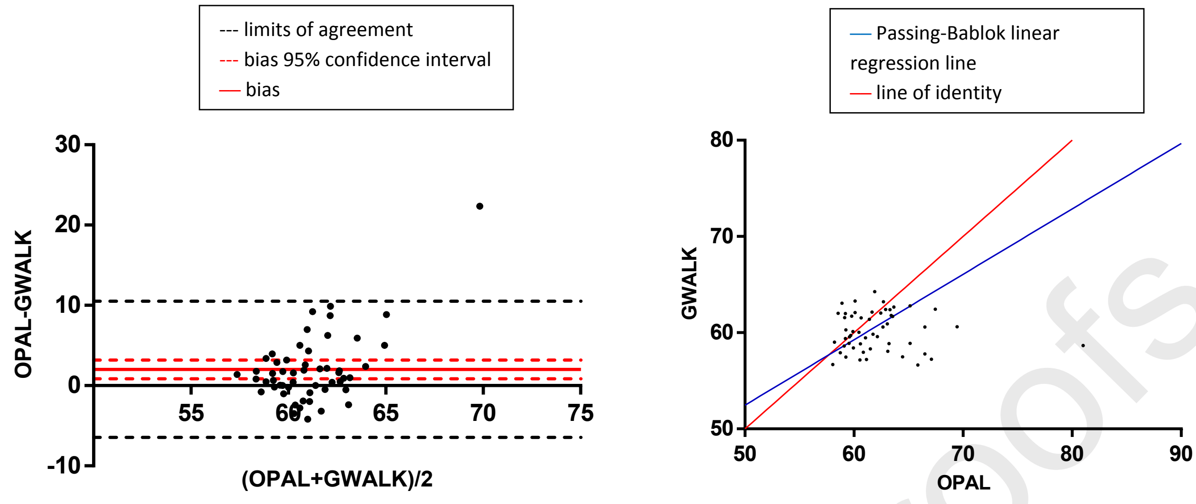

Figure 8 – Bland-Altman plot (difference vs average) of the STANCE PHASE parameter on the left - bias computed as the mean of differences, bias 95% confidence interval computed as $bias \pm 1.96\frac{SD}{\sqrt{N}}$, limits of agreement computed as $bias \pm 1.96\, SD$ where SD is the standard deviation of the differences and N is the sample size

Scatter plot of the STANCE PHASE parameter on the right - line of identity: GWALK = OPAL, Passing-Bablok linear regression line: linear regression line computed according the Passing-Bablok method

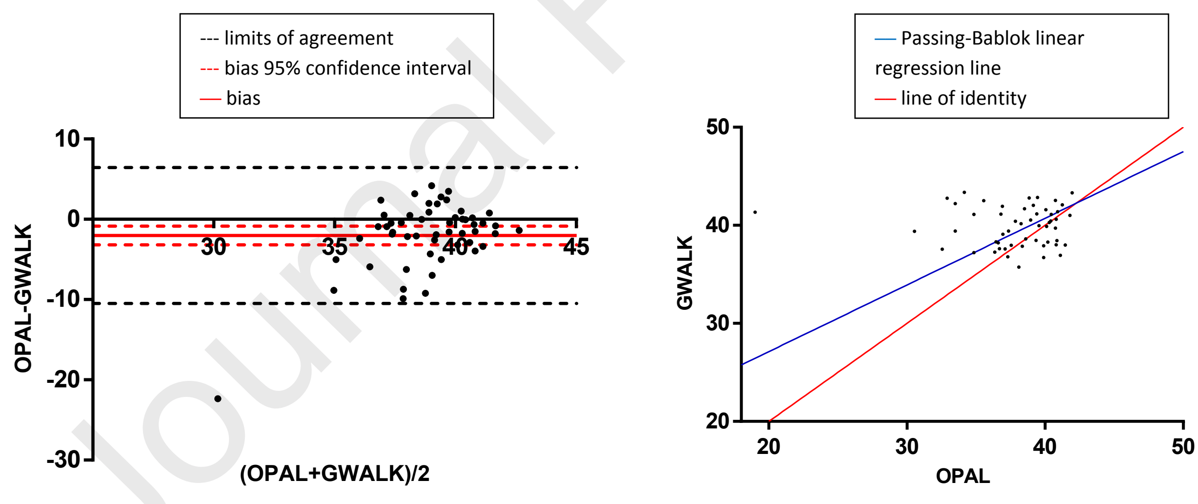

Figure 9 – Bland-Altman plot (difference vs average) of the SWING PHASE parameter on the left - bias computed as the mean of differences, bias 95% confidence interval computed as $bias \pm 1.96\frac{SD}{\sqrt{N}}$, limits of agreement computed as $bias \pm 1.96\, SD$ where SD is the standard deviation of the differences and N is the sample size

Scatter plot of the SWING PHASE parameter on the right - line of identity: GWALK = OPAL, Passing-Bablok linear regression line: linear regression line computed according the Passing-Bablok method

4 Discussion

The study evaluated the agreement and repeatability of spatiotemporal gait parameters computed by two wearable inertial systems for gait analysis: Opal System by APDM Inc. and G-Walk System by BTS Bioengineering Inc. so that clinicians and researchers can better interpret gait data obtained using these systems.

Concerning Cadence you can see a high Pearson's correlation coefficient between the measures carried out by the two systems ($r = 0.98$). Considering the Bland-Altman Analysis (Table 1) a bias equal to -1.59 and a 95% confidence interval (-2.31 ÷ -0.88) assures the presence of a statistical significant difference between the measures of Cadence computed by the two systems. Moreover the limit of agreement are quite wide (-6.80 ÷ 3.61). Figure 3 (Bland-Altman plot) shows that the vast majority of the points are under the line of zero (OPAL = GWALK), this means that Opal System underestimates cadence compared to G-Walk System, but their distribution around the line of bias is quite random, this assures the presence of only a constant systematic error statistically significant easy to eliminate through zeroing of bias. The results of Bland-Altman analysis are confirmed from Passing-Bablok regression analysis (Table 2), in fact the slope of the trend line is close to 1.00 ($m = 0.96$) and its 95% confidence interval (0.93 ÷ 1.00) includes the 1.00 value, this assures the absence of a proportional systematic error; moreover the intercept (offset) of the trend line is quite different from 0 ($q = 5.03$) and its 95% confidence interval (0.84 ÷ 9.22) does not include 0 value, this assures the presence of a constant systematic error. The scatter plot in Figure 3 shows that the vast majority of the points are over the line of identity (GWALK = OPAL), this means that G-Walk overestimates Cadence compared to Opal in according with the Bland-Altman plot. Finally Paired t-test analysis (Table 3) shows a highly significant overall difference ($p < 0.0001$) relative to Cadence parameter, G-Walk System shows an overall greater value than Opal System, this result is in according to the Bland-Altman and Passing-Bablok analysis. We can conclude that there is not agreement between Opal System and G-Walk System about Cadence parameter for the presence of a constant systematic error of about 5 steps/min that means difference of about 4.5 % between the two measurement systems.

Concerning Stride Velocity you can observe a high Pearson's correlation ($r = 0.89$). Considering the Bland-Altman Analysis (Table 1) a bias equal to 0.13 and a 95% confidence interval of the bias ranging from 0.09 to 0.16 assures the presence of a statistical significant difference. Moreover observing the graph in Figure 4 (Bland-Altman plot) we can observe not only that the vast majority of the points are over the line of zero (observed even in the scatter plot with an opposite situation due to the choice of the axis) that assures the presence of a constant systematic error (Opal overestimates stride velocity compared to G-Walk), confirmed even by Passing-Bablok regression analysis with an offset equal to -0.60 and a 95% confidence interval ranging from -0.90 and -0.32, but also a proportional error (negative trend in the data) with a moderate correlation ($r = -0.34$) between difference and mean of the Bland-Altman analysis, confirmed even by the Passing-Bablok regression with a slope equal to 1.36 and a 95% confidence interval ranging from 1.15 and 1.58. Paired t-test shows a highly overall difference ($p < 0.0001$) with Opal System that overestimates the measure of the Stride Velocity parameter, result that is in according to the results carried out from Bland-Altman analysis and Passing-Bablok regression analysis. We can conclude that there is not agreement between the two systems about Stride Velocity for the presence of a constant and a proportional systematic errors.

Concerning Gait Cycle Time we have a prefect agreement. There is a high Pearson's correlation ($r = 0.99$). From the Bland Altman analysis (Table 1 and Figure 5) a bias equal to 0, a 95% confidence interval of the bias close to 0, limits of agreement very tight and close to 0 too, a random distribution of the points around the line of zero assure the condition of a prefect agreement, confirmed even by Passing-Bablok regression with a slope close to 1 ($m = 0.98$) and its 95% confidence interval including 1, with an offset close to 0 ($q = 0.03$) and its 95% confidence interval including 0. Paired t test confirms the agreement with an overall difference not statistically significant ($p > 0.5$). Since Cadence and Gait Cycle Time are related, this result seems to be in disagreement with the previous one about Cadence. Actually, in the case of Cadence analysis, there is only a bias, a systematic error. The Cadence parameter should be calculated on a longer time and on a wider number of steps. The test protocol used in fact lasts only few seconds and the unit is steps/min, this condition can influence the two measurement systems in a different way determining a bias among them.

Repeating the same reasoning for Stride Length and Stride Length % that are linked, considering the Table 1, 2, 3 and the Figure 6, 7 we can conclude that there is not an agreement for the presence of a constant and proportional systematic errors. The constant error in fact is guaranteed by the presence of a bias statistically significant (Table 1) with the most of the points over the line of zero in the Bland-Altman plot and under the line of zero in the scatter plot (Figures 6, 7) and by the presence of an offset statistically significant (Table 2). The proportional systematic error is guaranteed by the presence of a negative trend in the Bland-Altman plot (Figures 6, 7) and by the presence of a slope statistically different from zero with a 95% confidence interval not including the zero value. Finally Paired t-test shows the presence of an overall statistical difference between the two measurement system about these two motion parameters with Opal System that overall overestimates the measures of the two parameters.

Both Opal and G-Walk Systems exhibited good agreement compared to two different gold standard in gait analysis on the parameters above analyzed [9,36] even if Opal System reached the best results. In the present study it emerges a poor agreement among them on these motion parameters.

For the Swing Phase and Stance Phase that are complementary, we can observe in Table 2 that the data computed by the two systems are strongly uncorrelated ($r = -0.05$), results confirmed graphically by the scatter plots in Figures 8, 9 where data assume a spherically symmetric, this means that there is not a linear dependence and therefore a strong correlation that we expected, even if it is a necessary but not sufficient condition. This result is enough to conclude that this is the worst not agreement evaluated. In fact since the two measurement systems are measuring the same parameter we expect that the point are on a line with slope equal to 1 and offset equal to 0. Anyway from the Bland-Altman analysis, Passing Bablock regression analysis and Paired T-test the presence of a not agreement emerged too. On these two parameters De Ridder et al. [36] found poor levels of agreement between G-Walk System and the gold standard GAITRite walkway System while excellent agreement for the other spatiotemporal parameters letting to suppose, considered our results too, the difficult of the G-Walk System to compute Swing Phase and Stance Phase, which are parameters based on final foot contact. On these two parameters, in fact, we had the worst results in terms of agreement. Another paper studied the agreement between G-Walk System and GAITRite System showing a bad agreement on these two parameters both on a healthy and pathological population [37]. Finally Bravi et al. also noticed incorrect gait cycle phase recognition by G-Walk [38]. For the Opal System Washabaugh et al. [9] found better results in terms of agreement with a gold standard on these two parameters with a Lin's concordance correlation coefficient (LCC) equal to 0.76. Based on our results and those provided by the scientific literature of reference, we can conclude that the strong not agreement between Opal and G-Walk Systems on these two motion parameters is attributable to G-Walk.

Repeatability through the calculation of ICC(3,1) shows important results. For the Opal System we can observe (Table 4) a ICC(3,1) value always greater than 0.95 for all spatiotemporal parameters analyzed, while for the G-Walk System we observe (Table 4) values minor than 0.95, except for the Gait Cycle Time parameter on which we had a perfect agreement between the two systems, but however high and greater than 0.8. We can conclude that Opal System has an excellent repeatability while G-Walk System a good repeatability about the study carried out. Repeatability study about G-Walk is in line with the study conducted by De Ridder et al. [36], they found an excellent test-retest reliability between consecutive measurements on the same day with intraclass correlation coefficient values ranging from 0.85 to 0.99. Our Test-retest reliability results of OPAL system are in line with the scientific literature, Washabaugh et al. [9] in fact found an excellent repeatability on spatiotemporal parameters with a ICC(3,1) ranging from 0.92 to 0.99.

It should be noted that the two measurement systems although based on the same technology (IMUs) and a different positioning of sensors, they present different hardware and software features that could influence the calculation of parameters. Anyway these three features: hardware, software and positioning of sensor are linked.

5 CONCLUSIONS

Study results suggest that gait analysis evaluation carried out by the two wearable inertial system for gait analysis based on the same technology (IMU sensors) but on a different sensor placement: G-Walk and Opal Systems does not give completely overlapping estimation of the spatiotemporal parameters analyzed, except for the Gait Cycle Time parameter, although they both showed good repeatability. This result discourages the use of the two systems in an interchangeable way. There have been two different types of error between the two methods of measurement: a constant systematic error (easily eliminable by zero bias by means of a simple subtraction operation from initial data) and a proportional systematic error. There is a need, within the community of the gait analysis laboratories, to understand which are the errors and disagreements that can be committed in the evaluation of kinematic parameters computed by different systems for gait analysis in order to use systems interchangeably for kinematic assessment of patients with motor disabilities. This kind of study is very important in the context of meta-analysis and where multi-centre studies are to be carried out to study specific pathologies and data coming from several different clinical facilities also acquired with different technology. Ensuring a high degree of agreement, specifically for each motion parameter, between different measurement systems for gait analysis and good repeatability guarantees to have reliable measurement if systems are used interchangeably and therefore follow the rehabilitation outcome of patients in a right way or perform research study with high quality. However, the clinical impact of a statistically significant error should be interpreted on the basis of established acceptability criteria: it is possible that relatively small proportional or constant systematic errors, although statistically significant, they do not preclude the use of a given method.

ACKNOWLEDGMENT

Authors thank Dr. F. Gallo of the Neurology Unit of ICS Maugeri Institute of Care and Scientific Research of Telese Terme (BN, Italy) for her strong support in patients' selection and assistance during the gait analysis sessions.

REFERENCES

[1] Sheng, B., Wang, X., Hou, M., Huang, J., Xiong, S., & Zhang, Y. (2020). An automated system for motor function assessment in stroke patients using motion sensing technology: a pilot study. *Measurement*, 107896. https://doi.org/10.1016/j.measurement.2020.107896

[2] Kamio, C., Aihara, T., & Minorikawa, G. (2020). Development of human movement measurement device for long-term measurements. *Measurement*, 107671. https://doi.org/10.1016/j.measurement.2020.107671

[3] Chen, J., Huang, Y., Guo, X., Zhou, S., & Jia, L. (2020). Parameter identification and adaptive compliant control of rehabilitation exoskeleton based on multiple sensors. *Measurement*, 107765. https://doi.org/10.1016/j.measurement.2020.107765

[4] Nonnekes, J., Goselink, R. J., Růžička, E., Fasano, A., Nutt, J. G., & Bloem, B. R. (2018). Neurological disorders of gait, balance and posture: a sign-based approach. *Nature Reviews Neurology*, *14*(3), 183.

[5] Eastlack, M. E., Arvidson, J., Snyder-Mackler, L., Danoff, J. V., & McGarvey, C. L. (1991). Interrater reliability of videotaped observational gait-analysis assessments. *Physical therapy*,*71*(6), 465-472. **https://doi.org/10.1093/ptj/71.6.465**

[6] Movement Disorder Society Task Force on Rating Scales for Parkinson's Disease. (2003). The unified Parkinson's disease rating scale (UPDRS): status and recommendations. *Movement Disorders*, *18*(7), 738-750. **https://doi.org/10.1002/mds.10473**

[7] Sharrack, B., & Hughes, R. A. (1996). Clinical scales for multiple sclerosis. *Journal of the neurological sciences*, *135*(1), 1-9. **https://doi.org/10.1016/0022-510X(95)00261-Y**

[8] Novacheck, T. F., Stout, J. L., & Tervo, R. (2000). Reliability and validity of the Gillette Functional Assessment Questionnaire as an outcome measure in children with walking disabilities. *Journal of pediatric orthopaedics*, *20*(1), 75.

[9] Washabaugh, E. P., Kalyanaraman, T., Adamczyk, P. G., Claflin, E. S., & Krishnan, C. (2017). Validity and repeatability of inertial measurement units for measuring gait parameters. *Gait & posture*, *55*, 87-93. **https://doi.org/10.1016/j.gaitpost.2017.04.013**

[10] Ong, A. M. L., Hillman, S. J., & Robb, J. E. (2008). Reliability and validity of the Edinburgh Visual Gait Score for cerebral palsy when used by inexperienced observers. *Gait & posture*,*28*(2), 323-326. **https://doi.org/10.1016/j.gaitpost.2008.01.008**

[11] Brunnekreef, J. J., Van Uden, C. J., van Moorsel, S., & Kooloos, J. G. (2005). Reliability of videotaped observational gait analysis in patients with orthopedic impairments. *BMC musculoskeletal disorders*, *6*(1), 17. **https://doi.org/10.1186/1471-2474-6-17**

[12] Viehweger, E., Pfund, L. Z., Hélix, M., Rohon, M. A., Jacquemier, M., Scavarda, D., ... & Simeoni, M. C. (2010). Influence of clinical and gait analysis experience on reliability of observational gait analysis (Edinburgh Gait Score Reliability).*Annals of physical and rehabilitation medicine*, *53*(9), 535-546. **https://doi.org/10.1016/j.rehab.2010.09.002**

[13] Lord, S. E., Halligan, P. W., & Wade, D. T. (1998). Visual gait analysis: the development of a clinical assessment and scale.*Clinical rehabilitation*, *12*(2), 107-119. **https://doi.org/10.1191/026921598666182531**

[14] D'Addio, G., Evangelista, S., Donisi, L., Biancardi, A., Andreozzi, E., Pagano, G., ... & Cesarelli, M. (2019, July). Development of a Prototype E-Textile Sock. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 17498-1752). IEEE. **DOI: 10.1109/EMBC.2019.8856739**

[15] Xu, X., McGorry, R. W., Chou, L. S., Lin, J. H., & Chang, C. C. (2015). Accuracy of the Microsoft Kinect™ for measuring gait parameters during treadmill walking. *Gait & posture*, *42*(2), 145-151. **https://doi.org/10.1016/j.gaitpost.2015.05.002**

[16] Agmon, M., Perry, C. K., Phelan, E., Demiris, G., & Nguyen, H. Q. (2011). A pilot study of Wii Fit exergames to improve balance in older adults. *Journal of geriatric physical therapy*, *34*(4), 161-167. **doi: 10.1519/JPT.0b013e3182191d98**

[17] Krishnan, C., Washabaugh, E. P., & Seetharaman, Y. (2015). A low cost real-time motion tracking approach using webcam technology. *Journal of biomechanics*, *48*(3), 544-548. **https://doi.org/10.1016/j.jbiomech.2014.11.048**

[18] Santos, T. M., Barroso, M. F., Ricco, R. A., Nepomuceno, E. G., Alvarenga, É. L., Penoni, Á. C., & Santos, A. F. (2019). A low-cost wireless system of inertial sensors to postural analysis during human movement. *Measurement*, *148*, 106933. https://doi.org/10.1016/j.measurement.2019.106933

[19] Wu, Y., Su, Y., Feng, R., Yu, N., & Zang, X. (2019). Wearable-sensor-based pre-impact fall detection system with a hierarchical classifier. *Measurement*, *140*, 283-292. https://doi.org/10.1016/j.measurement.2019.04.002

[20] Muro-De-La-Herran, A., Garcia-Zapirain, B., & Mendez-Zorrilla, A. (2014). Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications. *Sensors*, *14*(2), 3362-3394. https://doi.org/10.3390/s140203362

[21] D'Addio, G., Donisi, L., Mercogliano, L., Cesarelli, G., Bifulco, P., & Cesarelli, M. (2019, September). Potential Biomechanical Overload on Skeletal Muscle Structures in Students During Walk with Backpack. In *Mediterranean Conference on Medical and Biological Engineering and Computing* (pp. 262-266). Springer, Cham. **https://doi.org/10.1007/978-3-030-31635-8_31**

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

[23] Zheng, H., Black, N. D., & Harris, N. D. (2005). Position-sensing technologies for movement analysis in stroke rehabilitation.*Medical and biological engineering and computing*, *43*(4), 413-420. **https://doi.org/10.1007/BF02344720**

[24] Mancini, M., & Horak, F. B. (2016). Potential of APDM mobility lab for the monitoring of the progression of Parkinson's disease.*Expert review of medical devices*, *13*(5), 455-462. **https://doi.org/10.1586/17434440.2016.1153421**

[25] Di Leonardo, P., Beratto, L., Alessandria, M., ABATE DAGA, F., & Gollin, M. (2016). Core stability and unstable surface training in parkinson's desease: a longitudinal study. In *VIII Cong. Naz. SISMES, Ricerca e formazione applicate alle scienze motorie e sportive, 6-8 Ottobre. Roma 2016* (pp. 77-77).

[26] Bland, J. M., & Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet*, *327*(8476), 307-310. **https://doi.org/10.1016/S0140-6736(86)90837-8**

[27] Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical methods in medical research*, *8*(2), 135-160. **https://doi.org/10.1177/096228029900800204**

[28] Altman, D. G., & Bland, J. M. (1983). Measurement in medicine: the analysis of method comparison studies. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *32*(3), 307-317. **https://doi.org/10.2307/2987937**

[29] Sotgia, S., Zinellu, A., Pinna, G. A., Deiana, L., & Carru, C. (2008). A new general regression-based approach for method comparison studies. *Clinical chemistry and laboratory medicine*,*46*(7), 1046-1049. **https://doi.org/10.1515/CCLM.2008.194**

[30] D'Addio, G., Donisi, L., Pagano, G., Improta, G., Biancardi, A., & Cesarelli, M. (2019, July). Agreement between Opal and G-Walk Wearable Inertial Systems in Gait Analysis on Normal and Pathological Subjects. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 3286-3289). IEEE. **DOI: 10.1109/EMBC.2019.8857841**

[31] Passing, H., & Bablok, W. (1983). A new biometrical procedure for testing the equality of measurements from two different analytical methods. Application of linear regression procedures for

method comparison studies in clinical chemistry, Part I.*Clinical Chemistry and Laboratory Medicine*, *21*(11), 709-720. **https://doi.org/10.1515/cclm.1983.21.11.709**

[32] Blair, R. C., & Cole, S. R. (2002). Two-sided equivalence testing of the difference between two means. *Journal of Modern Applied Statistical Methods*, *1*(1), 18. **DOI: 10.22237/jmasm/1020255540**

[33] McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological methods*,*1*(1), 30. **https://doi.org/10.1037/1082-989X.1.1.30**

[34] Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*, *86*(2), 420. **https://doi.org/10.1037/0033-2909.86.2.420**

[35] Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, *15*(2), 155-163. **https://doi.org/10.1016/j.jcm.2016.02.012**

[36] De Ridder, R., Lebleu, J., Willems, T., De Blaiser, C., Detrembleur, C., & Roosen, P. (2019). Concurrent validity of a commercial wireless trunk triaxial accelerometer system for gait analysis. *Journal of Sport Rehabilitation*, *28*(6). https://doi.org/10.1123/jsr.2018-0295

[37] Vítečková, S., Horáková, H., Poláková, K., Krupička, R., Růžička, E., & Brožová, H. (2020). Agreement between the GAITRite® System and the Wearable Sensor BTS G-Walk® for measurement of gait parameters in healthy adults and Parkinson's disease patients. *PeerJ*, *8*, e8835. https://doi.org/10.7717/peerj.8835

[38] Bravi, M., Gallotta, E., Morrone, M., Maselli, M., Santacaterina, F., Toglia, R., ... & Miccinilli, S. (2020). Concurrent validity and inter trial reliability of a single inertial measurement unit for spatial-temporal gait parameter analysis in patients with recent total hip or total knee arthroplasty. *Gait & Posture*, *76*, 175-181. https://doi.org/10.1016/j.gaitpost.2019.12.014

Highlights

- Wearable inertial systems for gait analysis are spreading in clinical practice
- Opal and G-Walk Systems are based on a different sensor placement
- Validity and repeatability studies of measurement systems are of paramount importance
- Benchmarking analysis is fundamental for meta-analysis and multicenter studies
- Benchmarking was studied by Bland-Altman, Passing-Bablok regression, T-Test and ICC

Leandro Donisi: Conceptualization, Methodology, Formal Analysis, Writing - Original draft

Gaetano Pagano: Resources, Visualization

Giuseppe Cesarelli: Validation

Armando Coccia: Investigation

Federica Amitrano: Data Curation

Giovanni D'Addio: Supervision, Project administration, Writing – Review & Editing

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☒The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

No interests to declare