

Supplementary Material for: A Pilot Study to Assess Feasibility of Collecting and Transmitting Clinical Trial Data with Mobile Technologies

By Russell C, Ammour N , Wells T, Bonnet N, Kruse M, Tardat A, Erales C, Shook T, Kirkesseli S, Hovsepian L, Pretorius S

Contents

1	Statistical methods	1
1.1	Discussion of Analytical Approaches	2
1.1.1	Limits of Agreement	2
1.1.2	Correlated-Errors-in-Variables Model.....	3
1.1.3	Correlation Coefficient between Activity Watches	3
2	Usability Questionnaire Approach.....	3
2.1	Investigator After Scenario Questionnaire.....	3
2.2	Subject Device Usability Questionnaire	4
3	Usability Questionnaire Results.....	5
3.1	Investigator After Scenario Questionnaire (ASQ).....	5
3.1.1	Part 1	5
3.1.2	Part 2	7
3.2	Subject Device Usability Questionnaire (Part 2).....	10

1 Statistical methods

The pilot study was a non-interventional study, although treatment was administered during Part 1 according to the Sanofi study protocol to cohorts of subjects in a dose-finding manner. Each subject served as his/her own control in the comparison of devices.

Analyses are purely descriptive; there are no hypotheses specified, and the study was not powered to show equivalence between device pairs. Proposed analyses were pre-specified and in alignment with study objectives. Any *post hoc* analyses were clearly labelled.

Methodology for the statistical analysis of data collected from two or more devices intended to measure the same clinical parameter (e.g., blood pressure) or from two or more endpoints intended to support a clinical objective (e.g., total activity score from an actigraphy device and SF-36 total score) are emerging in the literature. Regulatory guidelines require the establishment of repeatability and agreement but do not recommend a specific statistical approach for continuous endpoints. Therefore, two statistical methods comparing the same measurements from pairs of devices (except the activity watches) and one method assessing correlation between endpoints from the activity watches will be implemented.

Note that repeatability of a device is estimated by within-subject variability, i.e., based on two or more measurements collected on the same subject using the same device under similar conditions, assuming independence of the measurements. Agreement is assessed by accuracy (absence of bias) and precision (variability) between devices.

1.1 Discussion of Analytical Approaches

Two scenarios are presented by the device data collected in this study:

- Method comparison (agreement) between two similar devices measuring the same clinical endpoint on the same scale, and
- Correlation between two devices collecting measurements on different scales.

This section will introduce published methodology for each approach. The methodology to be applied to data collected from each mobile technology (MT) and clinical standard device (CSD) pair during Part 1 and Part 2 is described below.

For methods comparisons, two cases were investigated for some devices:

- The underlying mean value is constant (e.g., weight collected on Days 1 to 3 is not expected to change as a result of study participation) which enables assessment of repeatability as well as agreement.
- The underlying mean value is non-constant (e.g., pulse oximeter readings during hyperventilation test) which is used to assess agreement only.

1.1.1 Limits of Agreement

Bland and Altman [1, 2, 3, 4] provided methodology to quantify repeatability and agreement for comparing measurement methods in the presence of repeated measures on each subject using Limits of Agreement (LoA). In general, LoA are determined from the differences between measurements from two devices for each subject and an estimate of variability. The 95% LoA are calculated as $\text{mean} \pm 1.96 \times SD$. These limits are expected to contain 95% of future individual differences between the two devices.

The methodology acknowledges that some disagreement between the methods will exist, but the question is, how much disagreement will cause a difference in clinical interpretation so that the two methods cannot be used interchangeably? Therefore, clinical judgment is required for interpretation of the resulting LoA. Table 3 in the article lists the clinical acceptance ranges for clinical endpoints, excluding activity.

Note that some endpoints have multiple levels of acceptability defined in the medical literature; in these cases, the LoA will be compared to the highest level. These ranges were also used for context for the repeatability coefficients.

1.1.2 Correlated-Errors-in-Variables Model

The correlated-errors-in-variables (CEIV) model proposed by Francq and Govaerts [5] provides an adjustment to the Bland/Altman approach by accounting for the correlation between the error terms of the difference in methods (y-axis) and the average of the two methods (x-axis) potentially resulting in tolerance intervals (TI) or prediction intervals (PI) that provide better coverage probability than the Bland/Altman LoA. The CEIV analysis focused on the approach of ‘practical equivalence’.

1.1.3 Correlation Coefficient between Activity Watches

Due to the internal calculation of activity for each watch, the two watches are not expected to show agreement; instead correlation between the two devices was of interest. Methods proposed by Roy [6] and Hamlett [7] were used to characterize activity as measured by the activity watches by estimating between- and within-subject correlation coefficients between the two devices. The between-subject correlation was calculated to characterize the relationship between total steps based on the Striiv compared to total activity counts based on the Actiwatch, i.e., how well one device predicts the outcome of the other device. The within-subject correlation was calculated to characterize the relationship of an increase in total steps based on the Striiv within an individual to an increase in total activity counts based on the Actiwatch. Since the relationship between the two watches for these endpoints was nonlinear, the correlation coefficient was not considered to be a valid summary statistic. A log transformation was also considered for these data.

2 Usability Questionnaire Approach

2.1 Investigator After Scenario Questionnaire

The After Scenario Questionnaire (ASQ) was filled out by the Investigators during Part 1 and Part 2. For each MT and prior to the first enrolled subject for each part of the study, Section 1 of the ASQ was filled out by each Investigator to assess ease of use of the device based on the Investigator’s personal use of the device to take his/her own measurements. Therefore, each Investigator evaluated their personal experience with each of the six devices one time: the spirometer and blood pressure monitor were tested and rated by the investigator during Part 1, and the other four devices were tested and rated by the investigator during Part 2.

Section 1 ratings were coded to numeric values as follows:

- Very dissatisfied = 1
- Dissatisfied = 2
- Neither satisfied nor dissatisfied = 3
- Satisfied = 4
- Very satisfied = 5

Ratings were summarized for each question and device. An overall score was calculated for each investigator based on the average of the ratings for the four questions. The average score across investigators was summarized.

Section 2 of the ASQ was completed by the Investigator to rate how well each subject was able to use each device based on amount of help needed. The assessment was comprised of one question per device and was completed once for each subject following the first use of each device after training.

Ratings were assigned as follows:

- They could not complete the device without help
- A lot of help
- A moderate amount of help
- A little help
- No help at all

Missing data were not imputed for any of the questions. Ratings for the amount of help needed were summarized for each device.

2.2 Subject Device Usability Questionnaire

The Device Usability Questionnaire (DUQ) is a modification of the Post Study System Usability Questionnaire developed by Lewis. It was completed by the subject during Part 2 only for each of the six MT. Section 1 only was completed on Day 1, and Sections 1, 2 and 3 were completed on Day 3.

Satisfaction ratings were coded as described for Section 1 of the ASQ. Agreement ratings were coded to numeric values as follows:

- Strongly disagree = 1
- Disagree = 2
- Neither agree nor disagree = 3
- Agree = 4
- Strongly Agree = 5

If the subject indicated 'Not applicable' the code was set to missing. Missing data were not imputed for any of the questions. The average of the five ratings from Section 1 on Day 1 were calculated. The usability evaluation score was calculated as the average for all questions from Sections 1, 2 and 3 collected on Day 3.

DUQ scores were summarized for each device combining data from Group 1 and Group 2. Summary statistics were presented for the following questions and average scores:

- DUQ Day 1 Section 1 individual questions
- DUQ Day 1 Section 1 average score
- DUQ Day 3 Sections 1, 2, 3 individual questions

- DUQ Day 3 Sections 1, 2, 3 average score

3 Usability Questionnaire Results

3.1 Investigator After Scenario Questionnaire (ASQ)

3.1.1 Part 1

Only 13 of the 25 subjects who were enrolled in this part were evaluated by the 5 investigators.

3.1.1.1 A&D BP Monitor UA-767PBT-Ci

Table 3-1 Summary of Investigator After Scenario Questionnaire - A&D BP Monitor UA-767PBT-Ci - Part 1

Device Question	Statistic	Total
SATISFACTION WITH:		
Information on use	M	5
Neither satisfied or dissatisfied	n (%)	1 (20.0)
Satisfied	n (%)	3 (60.0)
Very Satisfied	n (%)	1 (20.0)
Ease of use	M	5
Satisfied	n (%)	4 (80.0)
Very Satisfied	n (%)	1 (20.0)
Time to use	M	5
Neither satisfied or dissatisfied	n (%)	1 (20.0)
Satisfied	n (%)	3 (60.0)
Very Satisfied	n (%)	1 (20.0)
Overall	M	5
Neither satisfied or dissatisfied	n (%)	1 (20.0)
Satisfied	n (%)	3 (60.0)
Very Satisfied	n (%)	1 (20.0)
DEGREE OF HELP REQUIRED BY SUBJECT		
They could not complete without help	n (%)	0
A lot of help	n (%)	0
A moderate amount of help	n (%)	0
A little help	n (%)	0
No help at all	n (%)	13 (100.0)

M = number of investigators; N = number of subjects evaluated by investigator

3.1.1.2 *Vitalograph asma-1 BT*

Table 3-2 Summary of Investigator After Scenario Questionnaire - Vitalograph Asma-1 BT - Part 1

Device Question	Statistic	Total
SATISFACTION WITH:		
Information on use	M	5
Neither satisfied or dissatisfied	n (%)	1 (20.0)
Satisfied	n (%)	3 (60.0)
Very Satisfied	n (%)	1 (20.0)
Ease of use	M	5
Satisfied	n (%)	5 (100.0)
Time to use	M	5
Satisfied	n (%)	4 (80.0)
Very Satisfied	n (%)	1 (20.0)
Overall	M	5
Neither satisfied or dissatisfied	n (%)	1 (20.0)
Satisfied	n (%)	4 (80.0)
DEGREE OF HELP REQUIRED BY SUBJECT	N	13
A moderate amount of help	n (%)	1 (7.7)
A little help	n (%)	6 (46.2)
No help at all	n (%)	6 (46.2)

M = number of investigators; N = number of subjects evaluated by investigator

3.1.2 Part 2

The Investigator ASQ scores for Part 2 are summarized in Table 3-3, Table 3-4, Table 3-5 and Table 3-6.

3.1.2.1 *Entra MyGlucohealth*

Table 3-3 Summary of Investigator After Scenario Questionnaire – Entra MyGlucohealth - Part 2

Device Question	Statistic	Total
SATISFACTION WITH:		
Information on use	M	2
Neither satisfied or dissatisfied	n (%)	1 (50.0)
Very Satisfied	n (%)	1 (50.0)
Ease of use	M	2
Neither satisfied or dissatisfied	n (%)	1 (50.0)
Satisfied	n (%)	1 (50.0)
Time to use	M	1
Satisfied	n (%)	1 (100.0)
Overall	M	2
Neither satisfied or dissatisfied	n (%)	1 (50.0)
Satisfied	n (%)	1 (50.0)
DEGREE OF HELP REQUIRED BY SUBJECT	N	14
A little help	n (%)	9 (64.3)
No help at all	n (%)	5 (35.7)

M = number of investigators; N = number of subjects evaluated by investigator

3.1.2.2 *Nonin 9560*

Table 3-4 Summary of Investigator After Scenario Questionnaire – Nonin 9560 - Part 2

Device Question	Statistic	Total
SATISFACTION WITH:		
Information on use	M	2
Neither satisfied or dissatisfied	n (%)	1 (50.0)
Very Satisfied	n (%)	1 (50.0)
Ease of use	M	2
Satisfied	n (%)	1 (50.0)
Very Satisfied	n (%)	1 (50.0)
Time to use	M	2
Neither satisfied or dissatisfied	n (%)	1 (50.0)
Very Satisfied	n (%)	1 (50.0)
Overall	M	2
Neither satisfied or dissatisfied	n (%)	1 (50.0)
Very Satisfied	n (%)	1 (50.0)
DEGREE OF HELP REQUIRED BY SUBJECT	N	18
A little help	n (%)	3 (16.7)
No help at all	n (%)	15 (83.3)

M = number of investigators; N = number of subjects evaluated by investigator

3.1.2.3 A&D Weight Scale UC-351PBT –Ci

Table 3-5 Summary of Investigator After Scenario Questionnaire –A&D Weight Scale UC-351PBT –Ci - Part 2

Device Question	Statistic	Total
SATISFACTION WITH:		
Information on use	M	2
Satisfied	n (%)	1 (50.0)
Very Satisfied	n (%)	1 (50.0)
Ease of use	M	2
Satisfied	n (%)	1 (50.0)
Very Satisfied	n (%)	1 (50.0)
Time to use	M	2
Satisfied	n (%)	1 (50.0)
Very Satisfied	n (%)	1 (50.0)
Overall	M	2
Satisfied	n (%)	1 (50.0)
Very Satisfied	n (%)	1 (50.0)
DEGREE OF HELP REQUIRED BY SUBJECT	N	18
No help at all	n (%)	18 (100.0)

M = number of investigators; N = number of subjects evaluated by investigator

3.1.2.4 *Striiv Fusion*

Table 3-6 Summary of Investigator After Scenario Questionnaire –Striiv Fusion - Part 2

Device Question	Statistic	Total
SATISFACTION WITH:		
Information on use	M	2
Neither satisfied or dissatisfied	n (%)	1 (50.0)
Satisfied	n (%)	1 (50.0)
Ease of use	M	2
Neither satisfied or dissatisfied	n (%)	1 (50.0)
Satisfied	n (%)	1 (50.0)
Time to use	M	2
Neither satisfied or dissatisfied	n (%)	1 (50.0)
Satisfied	n (%)	1 (50.0)
Overall	M	2
Neither satisfied or dissatisfied	n (%)	2 (100.0)
DEGREE OF HELP REQUIRED BY SUBJECT	N	18
No help at all	n (%)	18 (100.0)

M = number of investigators; N = number of subjects evaluated by investigator

3.2 Subject Device Usability Questionnaire (Part 2)

Subjects evaluated the usability (“ease of use”) of each device using a modified questionnaire (DUQ) derived from the Post Study System Usability Questionnaire (PSSUQ) developed by Lewis [8]. The questionnaire comprised 3 sections with a total of 26 questions. Section 1 was completed on Days 1 and 3 of the extension study and sections 2 and 3 on Day 3.

Table 3-7 Summary of Subject Device Usability Questionnaire – Day 1 - Part 2

Section/Question	A&D BP Monitor UA-767PBT-Ci (n [%])	Vitalograph asma-1 BT (n [%])	Entra MyGlucohealth (n [%])	Nonin 9560 (n [%])	A&D weight scale UC-351PBT-Ci (n [%])	Striiv Fusion (n [%])
Easy to use device (N=22)						
Strongly disagree	1 (4.5)	0	0	0	0	0
Disagree	0	0	0	0	0	1 (4.5)
Neither agree nor disagree	3 (13.6)	1 (4.8)	1 (4.5)	0	0	0
Agree	4 (18.2)	6 (28.6)	12 (54.5)	5 (23.8)	3 (13.6)	7 (31.8)
Strongly agree	14 (63.6)	14 (66.7)	9 (40.9)	16 (76.2)	19 (86.4)	14 (63.6)
Easy to complete measurements (N=22)						
Strongly disagree	1 (4.5)	0	0	0	0	0
Disagree	0	0	1 (4.8)	0	0	1 (4.5)
Neither agree nor disagree	2 (9.1)	2 (9.1)	1 (4.8)	0	0	0
Agree	4 (18.2)	4 (18.2)	9 (42.9)	4 (18.2)	4 (18.2)	6 (27.3)
Strongly agree	15 (68.2)	16 (72.7)	10 (47.6)	18 (81.8)	18 (81.8)	15 (68.2)
Felt confident using device (N=22)						
Strongly disagree	0	0	0	0	0	0
Disagree	0	0	0	0	0	0
Neither agree nor disagree	2 (9.1)	1 (4.5)	2 (9.1)	1 (4.5)	0	0
Agree	4 (18.2)	4 (18.2)	6 (27.3)	3 (13.6)	3 (13.6)	7 (31.8)
Strongly agree	16 (72.7)	17 (77.3)	14 (63.6)	18 (81.8)	19 (86.4)	15 (68.2)
Clear when measurements complete (N=22)						
Strongly disagree	0	0	1 (4.5)	0	0	0
Disagree	0	1 (4.5)	0	1 (4.5)	0	0
Neither agree nor disagree	1 (4.5)	1 (4.5)	1 (4.5)	3 (13.6)	0	8 (36.4)
Agree	3 (13.6)	3 (13.6)	7 (31.8)	6 (27.3)	3 (13.6)	5 (22.7)
Strongly agree	18 (81.8)	17 (77.3)	13 (59.1)	12 (54.5)	19 (86.4)	9 (40.9)
Easy to submit measurements (N=22)						
Strongly disagree	0	0	0	0	0	0
Disagree	0	0	0	0	0	0
Neither agree nor disagree	0	3 (13.6)	2 (9.1)	0	0	0
Agree	0	8 (36.4)	8 (36.4)	1 (4.5)	1 (4.5)	0
Strongly agree	0	11 (50.0)	11 (50.0)	1 (4.5)	2 (9.1)	1 (4.5)
Not applicable	22 (100.0)	-	1 (4.5)	20 (90.9)	19 (86.4)	21 (95.5)

The DUQs for the questions relating to overall satisfaction and willingness to use the devices are summarized in Table 3-8.

Table 3-8 Summary of Subject's Overall Satisfaction and Willingness to use Devices – Day 3 - Part 2

Section/Question	A&D BP Monitor UA-767PBT-Ci (n [%])	Vitalograph asma-1 BT (n [%])	Entra MyGlucohealth (n [%])	Nonin 9560 (n [%])	A&D weight scale UC-351PBT-Ci (n [%])	Striiv Fusion (n [%])
Overall satisfaction with device (N=22)						
Very dissatisfied	0	0	0	0	0	0
Dissatisfied	1 (4.5)	1 (4.5)	1 (4.5)	0	0	1 (4.5)
Neither satisfied nor dissatisfied	0	2 (9.1)	4 (18.2)	0	0	4 (18.2)
Satisfied	6 (27.3)	7 (31.8)	6 (27.3)	5 (31.3)	3 (13.6)	8 (36.4)
Very satisfied	15 (68.2)	12 (54.5)	11 (50.0)	11 (68.8)	19 (86.4)	9 (40.9)
Willingness to use device >1/day for >1 month (N=22)						
Strongly disagree	2 (9.1)	1 (4.5)	2 (9.1)	0	1 (4.5)	3 (13.6)
Disagree	1 (4.5)	1 (4.5)	1 (4.5)	0	0	4 (18.2)
Neither agree nor disagree	1 (4.5)	1 (4.5)	3 (13.6)	0	0	3 (13.6)
Agree	4 (18.2)	5 (22.7)	9 (40.9)	5 (31.3)	6 (27.3)	5 (22.7)
Strongly agree	14 (63.6)	14 (63.6)	7 (31.8)	11 (68.8)	15 (68.2)	7 (31.8)

References

1. Altman DG, Bland JM. Measurement in Medicine: the Analysis of Method Comparison Studies. *The Statistician* 32 (1983) 307-317.
2. Bland JM, Altman DG. Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement. *Lancet*, 1986; i: 307-310.
3. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Statistical Methods in Medical Research* 1999; 8: 135-160.
4. Bland JM, Altman DG. Agreement Between Methods of Measurement with Multiple Observations per Individual. *Journal of Biopharmaceutical Statistics*, 17: 571-582, 2007.
5. Francq BG and Govaerts B. How to regress and predict in a Bland-Altman plot? Review and contribution based on tolerance intervals and correlated-errors-in-variables models. *Statistics in Medicine* 2016; 35:2328-2358.
6. Roy A. Estimating correlation coefficient between two variables with repeated observations using mixed effects model. *Biometrical Journal* 2006; 48(2):286-301.

7. Hamlett A, Ryan L, and Wolfinger R. On the use of PROC MIXED to estimate correlation in the presence of repeated measures. SUGI 29: Statistics and Data Analysis; Paper 198-29.
8. Lewis JR. Psychometric evaluation of the PSSUQ using data from five years of usability studies. *Int J Hum Comput Interact* 2002; 14(3-4):463–488.