Comparability of a Provisioned Device Versus Bring Your Own Device for Completion of Patient-Reported Outcome (PRO) Measures by Participants with Chronic Obstructive Pulmonary Disease (COPD): Quantitative Study Findings

Stacie Hudgens,[1] Louise Newton,[2] Sonya Eremenco,[3] Mabel Crescioni,[4] Tara Symonds,[2] Philip C.G. Griffiths,[2] Oliver Knight-West,[2] David S. Reasner,[5] Bill Byrom,[6] Paul O'Donohoe[7] and Susan Vallow[8] on behalf of the Patient-Reported Outcome (PRO) Consortium and the ePRO Consortium

[1]Clinical Outcomes Solutions, Tucson, AZ, US; [2]Clinical Outcomes Solutions, Folkestone, UK; [3]Patient-Reported Outcome (PRO) Consortium, Critical Path Institute, Tucson, AZ, US; [4]Electronic Patient-Reported Outcome (ePRO) Consortium, Critical Path Institute, Tucson, AZ, USA; [5]Ironwood Pharmaceuticals, Cambridge, MA, US; [6]CRF Bracket, London, UK; [7]Medidata Solutions, London, UK; [8]MedAvante-ProPhase, Hamilton, NJ, US

INTRODUCTION

- Historically, handheld devices have been provided to participants to report patient-reported outcome (PRO) data throughout a clinical trial (a provisioned device [PD]).
- The desire to reduce drug development costs and ease patient burden, combined with better access to handheld devices (e.g., smartphones and tablets), has led to increasing interest in having participants use their own devices ('bring your own device' [BYOD]) in clinical trials and other research studies.
- This approach now seems even more viable for three reasons:
  1. High data security is available for data storage and transfer for both PD and BYOD.
  2. A growing body of evidence supporting migration equivalence of paper to a variety of electronic formats supports the hypothesis that small format changes between devices may be inconsequential.[1]
  3. Technology allows for a similar display of items across device types.
- In previous studies assessing the feasibility of using a BYOD approach for PRO data collection, high compliance was found where most participants did not find the process of completing daily items burdensome.[2-5]
- Although promising in terms of the use of a BYOD approach, these studies did not aim at assessing BYOD in the context of a clinical trial and, more importantly, provided limited evidence for the feasibility of BYOD in a registrational clinical trial.

OBJECTIVE

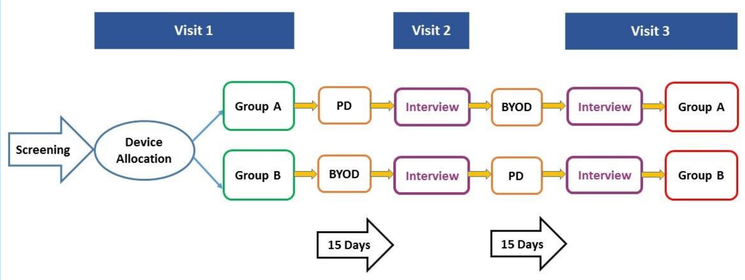
- To quantitatively compare PRO data collected via PD versus BYOD in terms of equivalence and compliance.

METHODS

- Participants with a clinical diagnosis of chronic obstructive pulmonary disease (COPD) were recruited for this observational, cross-over study conducted at four clinical sites in the US.
- Participants were asked to complete three study visits. Participants who met the inclusion criteria were allocated at Visit 1 to receive a PD (Group A) or instructed to use their own device (BYOD; Group B) for collection of PRO data. Group A were supplied with the software already installed on the PD and Group B downloaded and activated the application on their own device (BYOD).
- Both groups completed a training module on their respective device (Figure 1). Training included practice with completion of items from each PRO measure, information detailing when to complete the PRO measures, how to understand the recall period, how to choose the best option when unsure of the answer, and guidance that there was no "right" or "wrong" answer.

Methods - Continued

Figure 1. Study Design Flow Diagram



BYOD - bring your own device, **PD** - provisioned device

- For 15 days, participants completed the following PRO measures on either a PD or BYOD:
  - EXAcerbations of Chronic pulmonary disease Tool [EXACT®] (completed daily), and
  - COPD Assessment Test™ [CAT] and Patient Global Impression of Severity [PGIS]) (completed on Days 1, 8, and 15).
- At Visit 2, participants then switched to the other device type to complete the same measures for another 15 days.
- At Visit 3, a subset of 20 participants were invited to complete the screen size equivalence test during which the EXACT® and CAT were completed on a provisioned smartphone, tablet, and laptop at the site with a distraction task in between each completion.
- The EXACT® was scored using the Evaluating Respiratory Symptoms in COPD (E-RS: COPD™) algorithm.
- Daily and weekly compliance were evaluated descriptively.
- Equivalence was evaluated for all participants using:
  - Descriptive information between crossover periods
  - A linear mixed effects model, adjusting for cross-over sequence, period, and time (day/week) was used to derive the intraclass correlation coefficients (ICC)
  - Two one-sided test (TOST) analysis on both the unadjusted means and the adjusted means derived from the mixed effects model
    - For 10%, 20%, and 40% of total score tested as equivalence margins
- Equivalence was also assessed among three different screen sizes (laptop, tablet, smartphone [PD]) in an optional site-based task at the end of the study.

Methods - Continued

Analytic Approach

- For evaluation of the PRO measures, the cross-sectional population (CSP) and the longitudinal period population (LPP) were created.
- The CSP included participants with data at the specific analysis time point.
- The LPP was defined as participants with data at Day 1 and the respective analysis follow-up period:
  - The LPP was further delineated with two additional population flags for identifying stable participants for the evaluation of measurement equivalence:
    - Stable Population 1 (SP1) included participants who indicated no change from baseline to Day 30 on the PGIS.
    - Stable Population 2 (SP2) included participants who indicated minimal (|1|) to no change from baseline to Day 30 on the PGIS.

Compliance

- Compliance was evaluated and presented for the daily diary, by device, for each study week (Week 1 [Day 1–7], Week 2 [Day 8–14], Week 3 [Day 16–22] and Week 4 [Day 23–29]). The summary was presented as the number of participants with 0 to 7 missing diary days in each weekly period.
- Item-level compliance for the CAT was conducted for each key study timepoint (Day 1, 8, 15, 16, 23, and 30).

Equivalence

- Descriptive: Equivalence between cross-over periods was presented descriptively.
- For statistical equivalence testing, the LPP, SP1, and SP2 flags were included and were used to test the null hypothesis in the ICC and TOST analyses. These analyses were conducted at prespecified timepoints and using a full cross-over design.
  - Corrected ICCs were calculated using the LPP, SP1, and SP2 populations as an estimate of equivalence between devices on the CAT and E-RS™: COPD.
  - Statistics presented for the TOST analysis included both:
    - the difference between devices across all timepoints as per the cross-over design, and
    - the difference between devices at prespecified assessment points.
  - The mean difference, confidence interval (CI) (defined as 100[1−2α]%), and the effect size for the difference of means were calculated for all TOST analyses.
- Where ICCs and TOSTs were conducted using a full cross-over design, data from both Group A and Group B were assessed at all available time points.
- Cross-over analysis adjusted for device sequence (Group A or Group B), study period (Period 1 or Period 2), and assessment timepoint (the relevant PRO assessment timepoint).

Screen Size Assessment

- A separate assessment of the equivalence between differing screen sizes for both the E-RS™: COPD total score and CAT total score was conducted using:
  - ICCs and associated mean difference effect size estimates (and confidence bands).

RESULTS

Demographic and Clinical Characteristics

- Sixty-four participants were enrolled (mean age [SD]: 59.0 [10.55]; 65.6% female; 51.6% Black/African American) (Table 1).
- Group A and Group B were of similar mean [SD] ages (Group A = 57.5 years [11.33]; Group B = 59.8 years [10.13]) and the gender make-up was similar to the overall sample (Group A n = 14 female, 60.9%; Group B n = 28 female, 68.3%).
- Race distributions amongst the two allocation groups did not reflect the overall population. Group A participants were more likely to be White (n = 13, 56.5%) whereas Group B participants were more likely to be Black or African American (n = 27, 65.9%).
- The overall sample was diagnosed with COPD for an average of 7.2 years prior to enrollment, and half of the sample had no prior exacerbations (n = 32, 50.0%), with around a third of the sample reporting a single previous exacerbation (n = 20, 31.3%).

Table 1. Baseline Demographic and Clinical Information (FAS)

Demographic	Group A[a] PD 1st N = 23	Group B[b] BYOD 1st N = 41[c]	Overall N = 64
Age			
Mean (SD), years Min–Max	57.5 (11.33) 40–75	59.8 (10.13) 40–77	59.0 (10.55) 40–77
Gender			
Female	14 (60.9%)	28 (68.3%)	42 (65.6%)
Male	9 (39.1%)	13 (31.7%)	22 (34.4%)
Race			
White	13 (56.5%)	13 (31.7%)	26 (40.6%)
Black/African American	6 (26.1%)	27 (65.9%)	33 (51.6%)
Other	4 (17.4%)	1 (2.4%)	5 (7.8%)
Education			
Did not complete high school	0	5 (12.2%)	5 (7.8%)
High school diploma	8 (34.8%)	12 (29.3%)	20 (31.3%)
Some college or certificate program	6 (26.1%)	13 (31.7%)	19 (29.7%)
College or university degree	9 (39.1%)	8 (19.5%)	17 (26.6%)
Graduate degree	0	3 (7.3%)	3 (4.7%)
Work Status			
Employed Full-Time	15 (65.2%)	11 (26.8%)	26 (40.6%)
COPD Severity			
Very mild	0	1 (2.4%)	1 (1.6%)
Mild	0	13 (31.7%)	13 (20.3%)
Moderate	18 (78.3%)	18 (43.9%)	36 (56.3%)
Severe	5 (21.7%)	8 (19.5%)	13 (20.3%)
Very severe	0	1 (2.4%)	1 (1.6%)

BYOD = bring your own device, **FAS** = full analysis set, **PD** = provisioned device, **SD** = standard deviation
[a] Group A is defined as PD to BYOD
[b] Group B is defined as BYOD to PD
[c] Although every effort was made to ensure that sample sizes for both groups were equivalent, an issue occurred during recruitment where nine participants who were supposed to start on PD were unable to do so because of device issues and, instead, had to start on BYOD.

Quantitative Findings

Compliance

- For the first 15 days (period 1), almost all Group A participants (who were using PD) completed at least 5 diary days in Week 1 (95.2%; n=20), compared with 89.7% (n=35) of Group B participants (using BYOD) (Figure 2 and Figure 3).
  - Compliance in Week 2 followed a similar pattern but was slightly higher among participants in both groups.
  - One Group B participant in Week 1 and two Group B participants in Week 2 did not complete any of the 7 diary days.
- Following the device cross-over, in period 2:
  - 100% of Group A participants (now using a BYOD) completed at least 5 diary days in Weeks 3 and 4.
  - Compliance among Group B participants (now using a PD) was lower than Group A participants, with 76.9% (n=30) and 89.2% (n=34) completing at least 5 days of diary data during Weeks 3 and 4, respectively.
    - Three Group B participants in Week 3 and four Group B participants in Week 4 did not complete any of the 7 diary days.

Figure 2. Number of Missing Diary Days in Period 1 (Weeks 1 and 2)

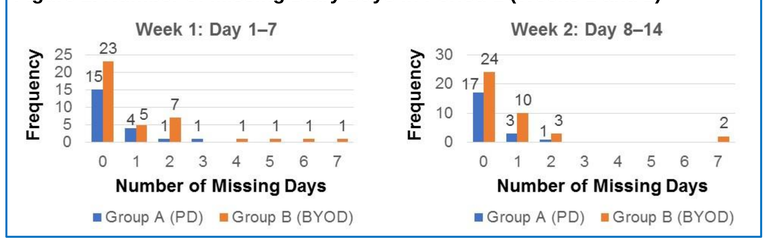
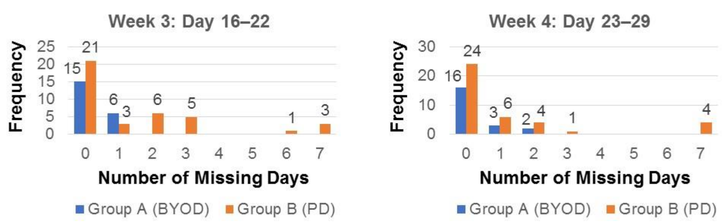


Figure 3. Number of Missing Days in Period 2 (Weeks 3 and 4)



Equivalence: Descriptive Assessment

- For each study week, participants' overall mean [SD] E-RS™: COPD score was similar (14.73[6.800]–15.28[6.497]), suggesting score equivalence between devices.
  - Table 2 shows the means for Group A and Group B independently
  - Means across cross-over period remain similar for both groups
- Overall, CAT scores between Period 1 and Period 2 were similar.
  - For Group A and Group B independently, means between the two cross-over periods showed minor differences (Group A: Day 1=20.7[8.88], Day 16=23.4[8.00]; Group B: Day 1=19.4[7.36], Day 16=18.2[8.51]).

Equivalence: ICC Based

- When scores were adjusted for order, study period, and assessment timepoint, in a full cross-over design using a mixed effects model, agreement across device types was high for the E-RS™: COPD weekly average score (ICC(2,1)=0.878) and the CAT score (ICC(2,1)=0.864).
  - For the SP1 population, consistency across device types was even higher for the E-RS™: COPD with ICC(2,1)=0.895. The SP2 population had high reliability but with the slightly lower ICC reflecting the increased variance expected with this sample; ICC(2,1)=0.863.
- Similarly, high agreement statistics were determined for the CAT when assessed in a full cross-over design using the SP1 (ICC(2,1)=0.908) and SP2 (ICC(2,1)=0.874) populations.
  - The CAT scores were also consistent between devices on Day 15 and Day 16, with ICC(2,1)=0.836.

Table 2. Score Level Descriptive Statistics for the Weekly E-RS™: COPD Stratified by Device Group (CSP)

Period	Visit	Statistic	Group A[a] (N = 21)	Group B[b] (N = 39)	Overall (N = 60)
Period 1	Week 1	n	21	35	56
		Mean (SD)	17.31 (5.961)	14.06 (6.580)	15.28 (6.497)
	Week 2	n	20	36	56
		Mean (SD)	16.96 (6.003)	13.58 (6.845)	14.79 (6.704)
Period 2 (device switch)	Week 3	n	21	32	53
		Mean (SD)	17.37 (5.755)	13.59 (6.456)	15.09 (6.410)
	Week 4	n	21	32	53
		Mean (SD)	17.57 (5.290)	12.87 (7.103)	14.73 (6.800)

BYOD = bring your own device; **CSP** = cross sectional population; **E-RS™: COPD** = Evaluating Respiratory Symptoms in COPD; **PD** = provisioned device
[a] Group A is defined as PD to BYOD
[b] Group B is defined as BYOD to PD

Equivalence: TOSTs

- When analyses were conducted using a full cross-over design, TOST analysis that used all available data and accounted for order, study period, and assessment timepoint, showing that the two device types were equivalent for both the E-RS™: COPD and the CAT scores.
  - Ninety percent (90%) confidence intervals around the differences between scores on the two device types fell within the ±10% equivalence levels
  - This result shows that both measures were equivalent within a ±10% (4-point) equivalence margin.
- TOSTs for CAT at Day 1, Day 8, Day 15, Day 16, Day 23, and Day 30 showed that the daily scores were equivalent between PD and BYOD at the 20% equivalence levels at most time points and the 40% equivalence for all time points.

Screen Size Equivalence

- Both the E-RS™: COPD and CAT measures performed consistently across the three device screen sizes, with high ICCs reported for both the E-RS™: COPD (ICC(2,1)=0.989) and the CAT (ICC(2,1)=0.972).

Limitations

- Device type assignment was not randomized due to limited number of PDs available, which led to differences between Group A and Group B that impacted compliance. Therefore, the study was not able to answer the question of whether compliance is better with PD or BYOD because of many confounding factors.
- In the second part of this cross-over study, there was an increase in the number of participants missing the entire week in Group B after switching to PD.
- Subgroup analyses revealed a possible order effect.
  - Participants preferred the device they used in the second study period.
  - Some planned analyses were not conducted due to this order effect.

CONCLUSIONS

- Diary completion was very high for both devices when assessing daily compliance as well as a threshold for the weekly assessment of five or more days of complete diary data.
- A key aim of this study was to assess equivalence between the two device types. Descriptively, the overall mean scores for the E-RS™: COPD and the CAT before and after the cross-over period were very similar, though means for Group B were slightly lower, reflecting a less severe population.
- Equivalence was demonstrated across both the PD and BYOD devices with confidence intervals within standard ranges for equivalence.
- This study supports the use of BYOD as a potential complement to PD for collecting PRO data in COPD studies and contributes evidence that BYOD may be employed to collect PRO data in diverse patient populations.
- Although non-compliance with PD occurred, this could be related to unequal sample sizes between the two groups.

ACKNOWLEDGMENTS

- Funding for this project was provided by Critical Path Institute's Patient-Reported Outcome (PRO) Consortium and Electronic Patient-Reported Outcome (ePRO) Consortium along with additional contributions from the following PRO Consortium member firms: Amgen Inc.; Bayer Pharma AG; Daiichi Sankyo, Inc.; GlaxoSmithKline; Ironwood Pharmaceuticals, Inc.; Janssen Pharmaceutical Companies of Johnson and Johnson; Eli Lilly and Company; Merck Sharp & Dohme Corp.; Sanofi; and, Pfizer, Inc. Critical Path Institute and the PRO Consortium are supported, in part, by Critical Path Public-Private Partnerships Grant Number U18 FD005320 (effective 2015-2020) from the U.S. Food and Drug Administration. Additional support for the PRO Consortium comes from membership fees paid by members of the PRO Consortium (https://c-path.org/programs/pro/). Support for the ePRO Consortium comes from membership fees paid by members of the ePRO Consortium (https://c-path.org/programs/epro/). The authors gratefully acknowledge the other members of the BYOD Project Team for their guidance and constructive feedback during the conduct of this research.

REFERENCES

1. Byrom B, Gwaltney C, Slagle A, et al. Measurement equivalence of patient-reported outcome measures migrated to electronic formats: a review of evidence and recommendations for clinical trials and bring your own device. *Ther Innov Regul Sci* 2018, doi: 10.1177/2168479018793369 (Epub ahead of print).
2. Michaud K, Schumacher R, Wahba K, Moturu S. Are rheumatic disease patient reported outcomes collected passively and directly through smart phones feasible? Early results from a nation-wide pilot study (FRI0201). *Ann Rheum Dis* 2014;73(Suppl 2):455-456.
3. Pfaeffli L, Maddison R, Jiang Y, Dalleck L. Measuring physical activity in a cardiac rehabilitation population using a smartphone-based questionnaire. *J Medical Internet Res* 2013;15(3):e61.
4. Torous J, Staples P, Shanahan M, et al. Utilizing a personal smartphone custom app to assess the patient health questionnaire-9 (PHQ-9) depressive symptoms in patients with major depressive disorder. *JMIR Mental Health* 2015;2(1):e8.
5. Byrom B, Doll H, Muehlhausen W, et al. Measurement equivalence of patient-reported outcome measure response scale types collected using bring your own device compared to paper and a provisioned device: results of a randomized equivalence trial. *Value Health* 2018;21:581-589.