




Home monitoring improves endpoint efficiency in idiopathic pulmonary fibrosis

Kerri A. Johannson^{1,2}, Eric Vittinghoff³, Julie Morisset⁴, Joyce S. Lee⁵, John R. Balmes^{6,7} and Harold R. Collard⁷

Affiliations: ¹Department of Medicine, University of Calgary, Calgary, AB, Canada. ²Department of Community Health Sciences, University of Calgary, Calgary, AB, Canada. ³Department of Biostatistics and Epidemiology, University of California San Francisco, San Francisco, CA, USA. ⁴Département de Médecine, Centre Hospitalier de l'Université de Montréal, Montréal, QC, Canada. ⁵Department of Medicine, University of Colorado Denver, Aurora, CO, USA. ⁶Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, CA, USA. ⁷Department of Medicine, University of California, San Francisco, CA, USA.

Correspondence: Kerri A. Johannson, 4448 Front Street S.E., Calgary, AB, Canada, T3M-1M4.
E-mail: kerri.johannson@ahs.ca

 @ERSpublications
Home monitoring improves endpoint efficiency in idiopathic pulmonary fibrosis #AdvancesinILD
@ERSTalk <http://ow.ly/F5L930bUypH>

Cite this article as: Johannson KA, Vittinghoff E, Morisset J, *et al.* Home monitoring improves endpoint efficiency in idiopathic pulmonary fibrosis. *Eur Respir J* 2017; 50: 1602406 [https://doi.org/10.1183/13993003.02406-2016].

ABSTRACT The objective of this study was to investigate the reliability, feasibility and analytical impact of home-based measurement of forced vital capacity (FVC) and dyspnoea as clinical endpoints in idiopathic pulmonary fibrosis (IPF).

Patients with IPF performed weekly home-based assessment of FVC and dyspnoea using a mobile hand-held spirometer and self-administered dyspnoea questionnaires. Weekly variability in FVC and dyspnoea was estimated, and sample sizes were simulated for a hypothetical 24-week clinical trial using either traditional office-based interval measurement or mobile weekly assessment.

In total, 25 patients were enrolled. Mean adherence to weekly assessments over 24 weeks was greater than 90%. Compared with change assessment using baseline and 24-week measurements only, weekly assessment of FVC resulted in enhanced precision and power. For example, a hypothetical 24-week clinical trial with FVC as the primary endpoint would require 951 patients using weekly home spirometry compared with 3840 patients using office spirometry measures at weeks 1 and 24 only. The ability of repeated measures to reduce clinical trial sample size was influenced by the correlation structure of the data.

Home monitoring can improve the precision of endpoint assessments, allowing for greater efficiency in clinical trials of therapeutics for IPF.

This article has supplementary material available from erj.ersjournals.com

Received: Dec 07 2016 | Accepted after revision: March 21 2017

Support statement: This study was supported by The Pulmonary Fibrosis Foundation and The CHEST Foundation Clinical Research Grant in Pulmonary Fibrosis. Funding information for this article has been deposited with the Crossref Funder Registry.

Conflict of interest: Disclosures can be found alongside this article at erj.ersjournals.com

Copyright ©ERS 2017

Introduction

Idiopathic pulmonary fibrosis (IPF) is a progressive diffuse parenchymal lung disease of complex aetiology [1–3]. Recently, two therapies (nintedanib and pirfenidone) that slow disease progression were approved for IPF, and most patients are now being offered these medications [4, 5]. This has complicated the clinical development pathway for novel therapies, most directly by slowing the rate of change in clinical endpoints and increasing the number of patients required to adequately power between-group comparisons.

Clinical endpoints in IPF trials have traditionally been measured every 3–4 months in an office-based setting by trained study coordinators. Forced vital capacity (FVC) remains the primary endpoint of choice in IPF clinical trials [6], and symptom severity (e.g. dyspnoea), quality of life and survival time are routinely measured as secondary endpoints [7]. More frequent measurement of clinical endpoints could improve analytical precision and reduce sample size requirements for future clinical trials [8], but requiring study subjects to return to the study centre weekly (or even monthly) has been considered a significant barrier to enrolment.

Mobile health tools now exist to allow for home assessment of many clinical variables, including FVC and symptom severity, but have yet to be adopted in clinical trials. Significant concerns relate to feasibility and data integrity, in particular when considering primary endpoints. However, recent data using hand-held spirometers to record daily FVC support the feasibility of home-based FVC assessment [9]. The objective of the current study was to further investigate the reliability, feasibility and analytical impact of home-based measurement of clinical endpoints in IPF, specifically FVC and dyspnoea. Secondary objectives were to characterise the short-term variability in FVC and dyspnoea measurements, and determine the correlation of change in FVC with change in dyspnoea.

Methods

Study population

Patients were prospectively recruited from the interstitial lung disease (ILD) programme at the University of California, San Francisco (UCSF). Eligibility criteria included having a diagnosis of IPF (according to current consensus guidelines) [1], residing in the state of California and not concomitantly participating in a treatment trial of IPF therapy. The study was approved by the UCSF Institutional Review Board, and all patients provided written informed consent.

Study design

Patients were enrolled in a longitudinal cohort study lasting up to 40 weeks, involving a baseline visit followed by remote measurement of key clinical variables. At the baseline visit, each patient's age, sex and smoking history were recorded. Office-based spirometry to American Thoracic Society standards [10] was also performed, and the University of California, San Diego shortness of breath questionnaire (UCSD-SOBQ) was completed [11]. Each patient was provided with a personalised hand-held spirometer (PMD Healthcare, Allentown, PA, USA) and received instruction in its use. The spirometer provided real-time feedback to the patient to ensure proper technique and to optimise compliance.

Hand-held spirometry was performed at the baseline visit and weekly at home for the duration of study involvement. Three FVC manoeuvres were performed at each assessment, with the highest value recorded by the spirometer. Prior to each use of the hand-held spirometer, patients completed the UCSD-SOBQ and a 10-point visual analogue scale (VAS) for dyspnoea severity (figure S1). Patients maintained a weekly diary documenting any unscheduled medical visits, any acute worsening of their respiratory symptoms and any changes to their medications over the prior week.

Statistical analysis

Means, proportions and standard deviations were used as appropriate to describe the study population. Adherence to home measurement was determined by the number of home measurements divided by the number of weeks enrolled in the study. Regression was used to identify factors associated with adherence, with candidate variables including baseline age, sex, lung function and symptom scores.

Change over 24 weeks in outcome variables was calculated by two methods: a between-group comparison of the change between the baseline and 24-week home measurements (method 1), and linear slopes estimated using linear mixed models based on all available weekly measurements (method 2). For method 2, week of measurement was included in the model. Baseline office-based and hand-held spirometry measurements were compared using the Bland–Altman method [12]. Similarly, the change in FVC from baseline to end of study, as measured by both office-based and hand-held spirometry, was compared.

Week-to-week variability in FVC, UCSD-SOBQ score and dyspnoea VAS was estimated using two measures: the subject-specific coefficient of variation, defined as the ratio of the standard deviation (SD) to the mean of all home measurements for each subject, and the absolute change in each measure from one week to the next. The relationships between the repeated week-to-week changes in FVC, UCSD-SOBQ score and dyspnoea VAS, both with each other and with age, sex, baseline lung function and symptoms scores, were estimated using linear mixed models.

Estimated sample sizes required for 24-week clinical trials using FVC, UCSD-SOBQ and VAS as primary endpoints were calculated for two contrasting approaches. In the first approach, ANCOVA was used to estimate the effect of treatment on change from baseline to final measurement, adjusting for the baseline value. In the second approach, linear mixed models were used to estimate the effect of treatment on the linear trend in the weekly home measurements; this model included terms for time, treatment and the interaction between them, and accounted for the within-subject correlation of the repeated outcomes. Type 1 error was set at 0.05. Power was set at 80%, and the between-group difference (effect size) was varied from 20% to 50%. The background 24-week “control group” change was assumed to be a loss of 0.05 L in FVC, a one-point decline in dyspnoea VAS, and an increase of six points in the UCSD-SOBQ (based on historical observations) [5, 13–15]. We performed similar calculations assuming an FVC decline of 0.10 L and 0.15 L over 24 weeks. Study data were used to estimate the standard deviation and within-subject correlation of the outcomes. For the mixed models, pairwise correlations between outcomes obtained 1, 2, 3, ..., 23, and 24 weeks apart were used to determine the correlation structure for each outcome. For the ANCOVA analyses, all participants were assumed to contribute at least one follow-up measurement. For mixed-model analysis, loss to follow-up was estimated at 10%, occurring steadily across the 24 weeks of the analysis. Analyses were performed using Stata v.14 (StataCorp, College Station, TX, USA) and R Version 3.3.1 (R Foundation for Statistical Computing, Vienna, Austria).

Results

Study population

In total, 30 patients were screened for enrolment between January 1, 2014 and September 30, 2014. Of these, four failed screening because of difficulties with use of the home spirometer and one withdrew prior to enrolment because of progressive disease. The remaining 25 patients constituted the study cohort. Baseline characteristics of the cohort are presented in table 1. The mean age was 74 years (SD=7.5), 84% were male, the mean absolute FVC was 2.7 L (SD=0.69), and mean FVC percent predicted was 68% (SD=15). Five patients were taking pirfenidone at study enrolment with an additional nine patients initiating therapy during the study period.

TABLE 1 Baseline patient characteristics

Subjects n	25
Age years	73.6±7.5
Male sex	21 (84)
Smoking status	
Never	8 (32)
Former	17 (68)
FVC absolute L	2.7±0.69
FVC % pred	68±15
FEV₁ absolute L	2.2±0.55
FEV₁ % pred	76±16
FEV₁/FVC ratio	0.81±0.06
UCSD-SOBQ score	36.3±29.5
VAS score	2.76±2.18
Surgical lung biopsy	10 (40)
Weeks enrolled	33.0±9.1
Change in FVC over 24 weeks L	-0.1±0.11
Change in FVC over 40 weeks L	-0.17±0.18
Change in UCSD-SOBQ over 24 weeks	0.18±13
Change in UCSD-SOBQ over 40 weeks	0.30±22
Change in VAS over 24 weeks	0.65±0.96
Change in VAS over 40 weeks	1.1±1.6

Data are presented as mean±SD or n (%), unless otherwise stated. FVC: forced vital capacity; % pred: % predicted; FEV₁: forced expiratory volume in 1 s; UCSD-SOBQ: University of California, San Diego Shortness of Breath Questionnaire; VAS: visual analogue scale.

Reliability and feasibility of home-based assessment

The baseline office-based and hand-held spirometer FVC values were highly correlated ($r=0.91$). In the subset of patients (15/25) with contemporaneous office-based and hand-held spirometry after study completion, correlation remained high ($r=0.96$). A Bland–Altman plot comparing office and home FVC show good overall agreement (figure 1), with similarly good agreement between change in FVC over time, comparing both techniques (figure S2). In total, 782 home measures of FVC, 827 home measures of UCSD-SOBQ and 820 home measures of VAS were recorded during the entire study period. Mean adherence to home monitoring over 24 weeks was 90.5% ($SD=18.3$; figure 2a). Three patients collected less than 50% of the scheduled home readings because of difficulties with the spirometer ($n=2$) or confusion about the study protocol ($n=1$). Adherence to home monitoring decreased over time, in particular after 24 weeks (figure 2b). There were no statistically significant relationships between adherence and baseline demographics, lung function or dyspnoea scores.

Analytical efficiency of home-based assessment

Weekly home-based assessment of FVC, UCSD-SOBQ, and dyspnoea VAS have discordant impacts on estimated sample size requirements for hypothetical 24-week clinical trials (table 2 and table S1). The impact of home-based FVC assessment is profound; in our hypothetical clinical trial, weekly measurement of FVC reduced the estimated sample size requirement by 75% (from 3840 to 951), for 80% power to detect an effect size of 50%, compared with intermittent measures taken solely at weeks 1 and 24. Home-based assessment of dyspnoea, however, had no meaningful impact on study power. This difference is explained by the correlation structure of the weekly home-based data. Weekly home FVC measurement demonstrated a stable correlation structure (*i.e.* the within-patient correlation between any pair of measurements did not depend on the time interval between them), whereas weekly dyspnoea assessment demonstrated a declining correlation structure (*i.e.* the within-subject correlation was lower for measurements obtained farther apart). Specifically, we found the within-subject correlation for FVC measurements was constant at 0.92, whereas the correlation for dyspnoea VAS declined from 0.92 to 0.69, and the correlation for UCSD-SOBQ declined from 0.98 to 0.85. The reduced correlation between measures as the time between them increased has important implications for study power and sample size requirements (see table 2, figure 4, online supplement and table S1).

Weekly variability and correlation in home-based assessments

The mean and median weekly variability in FVC was -10.5 mL ($SD=211$) and -10 mL (range -1210 to 1150), respectively, with a coefficient of variation of 8.2% ($SD=3.6$) (figure 3). The mean and median weekly variability in UCSD-SOBQ was 0.38 ($SD=4.9$) and 0 (range -24 to 24), respectively, with a coefficient of variation of 7.8% ($SD=3.5$), while for the dyspnoea VAS, the mean and median variability was 0.03 (0.77) and 0 (range -4 to 4), respectively, with a coefficient of variation of 39.1% ($SD=53$) (figure S3). Weekly variability in FVC was independent of weekly variability in UCSD-SOBQ or VAS scores with standardised coefficients of -0.06% (95% CI: -0.13 , 0.01 ; $p=0.07$) and -0.01 (95% CI: -0.07 , 0.06 ; $p=0.89$), respectively. The weekly variability in UCSD-SOBQ and VAS were highly correlated, with a standardised coefficient of 0.32% (95% CI: 0.26 , 0.38 ; $p<0.0001$).

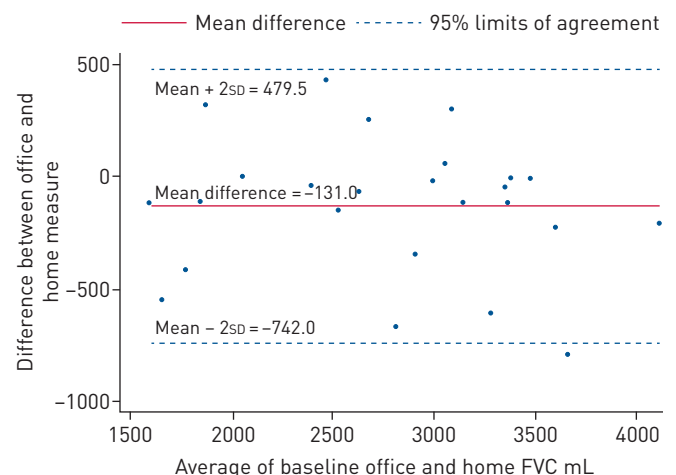


FIGURE 1 Bland–Altman plot comparing baseline office-based and hand-held spirometry. FVC: forced vital capacity.

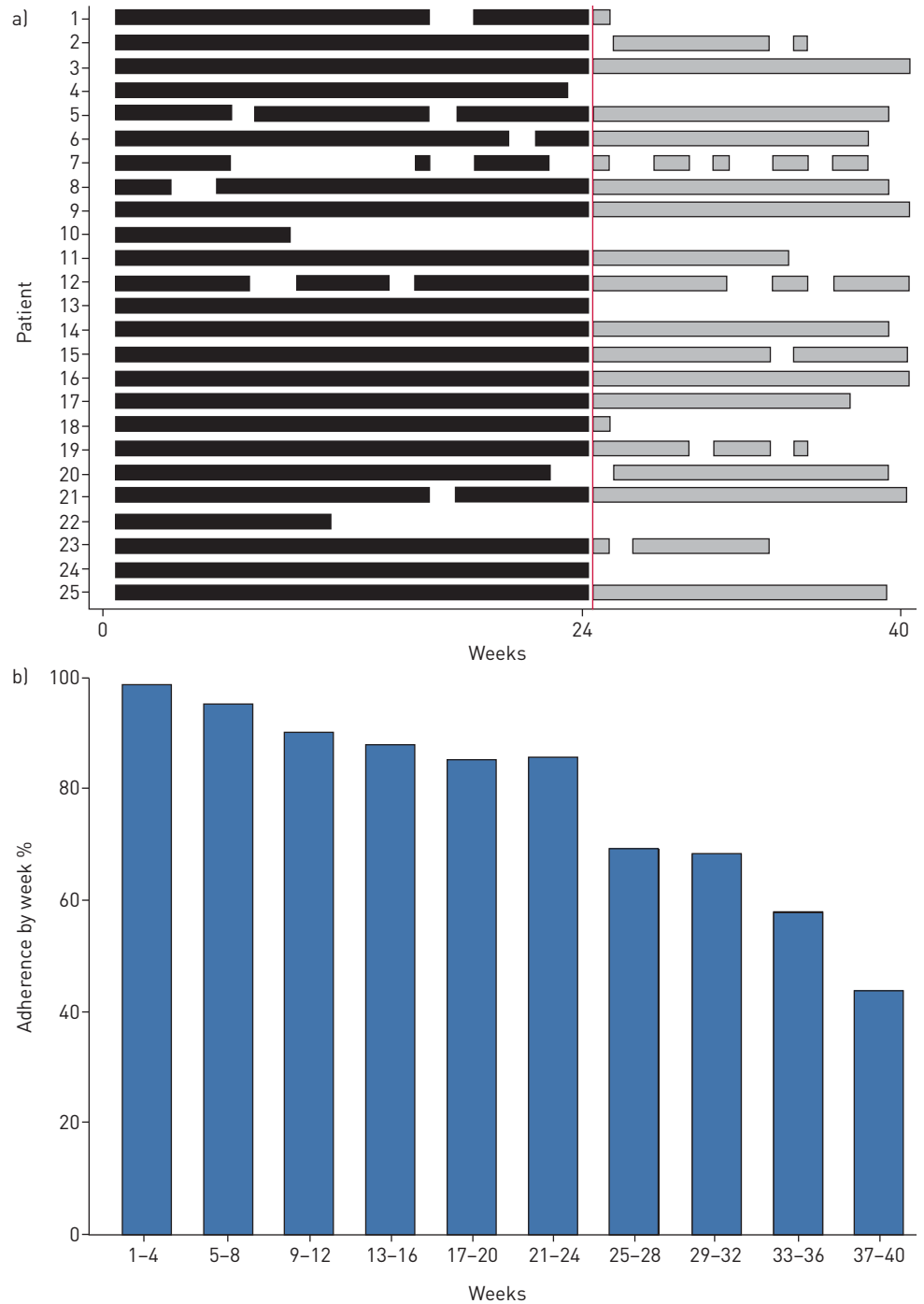


FIGURE 2 Adherence to home spirometry. a) Mean adherence over the study period. Black lines indicate measurements recorded between weeks 1 and 24, grey lines indicate measurements recorded between weeks 24 and 40. b) Percentage of patients adherent over the study period.

Discussion

Our results demonstrate that weekly home measurement of FVC and dyspnoea in patients with IPF is reliable and feasible over the course of 24 weeks, but that the ability of repeated measures to improve endpoint efficiency depends on certain statistical features of the data. Our data suggest that weekly home-based measurements of FVC have the potential to improve the efficiency of clinical trials while mitigating issues of missing data [16], but this may not be true in all circumstances, as illustrated by measures of dyspnoea. We believe our results validate the experience of RUSSELL *et al.* [9], and demonstrate home-based assessment to be a reliable and feasible method of endpoint measurement in most patients with IPF.

TABLE 2 Sample size estimates to achieve 80% power, comparing intermittent and repeated measures

Outcome	Effect size %	Measurement frequency	
		Weekly×24	Weeks 1 and 24
FVC; assumed control change of –50 mL	20	5946	24 002
	35	1942	7837
	50	951	3840
FVC; assumed control change of –100 mL	20	1487	6000
	35	485	1959
	50	238	960
FVC; assumed control change of –150 mL	20	661	2667
	35	216	871
	50	106	427
UCSD-SOBQ	20	1778	1731
	35	581	565
	50	285	277
Dyspnoea VAS	20	5122	4993
	35	1672	1630
	50	819	799

FVC: forced vital capacity; UCSD-SOBQ: University of California, San Diego Shortness of Breath Questionnaire; VAS: visual analogue score.

We had four patients (13%) who were unable to use the home spirometer, but the vast majority found it easy to manage and demonstrated excellent compliance for 24 weeks. Compliance worsened with prolonged use, and this may be an important issue when considering trials of longer duration. Compliance with home dyspnoea assessment was similarly excellent. No clear predictors of reduced compliance were found in this cohort. Data on reasons for discontinuation were not systematically collected. Blinding of study subjects to data generated by home measurement tools is possible and may be advisable, particularly in randomised blinded trials of therapeutics. Concerns over witnessed lung function decline may or may not prompt study patients to withdraw because of concerns over their allocated treatment arm. Real-time wireless uploading of spirometry data and other outcome measurements to a central portal that can be monitored may be an important tool to improve compliance, as adherence to home testing can be monitored in real time. It is likely that larger numbers of patients with varying levels of compliance will be needed to address this issue more thoroughly.

The relevance of data correlation structure to the analytical efficiency of home measurement is an unanticipated and important finding that deserves careful consideration. A common assumption of repeated measures analysis is that the within-subject measurement correlation is exchangeable, but as our data demonstrate, this is not always the case. In data with exchangeable correlation, the within-subject correlation between pairs of measurements does not depend on the time interval between them. With declining correlation, the within-subject correlation is lower between measurements obtained further apart in time. In practical terms, between-group differences in trend are relatively easy to detect in repeated measures data with exchangeable correlation. By contrast, weekly repeated measures data with declining correlation may provide little more information about between-group differences in trend than do pairs of measurements obtained at the beginning and end of the study. This implies that investigators planning a trial will need reliable pilot information about the correlation structure of the outcome in determining the potential utility of obtaining weekly home outcome measurements. We have provided sample size calculations for illustrative purposes to highlight the importance of correlation structure, and these numbers are not intended for clinical trial planning purposes. Careful work should be undertaken to better characterise the role of home monitoring to measure specific outcomes in more diverse populations.

Another key and cautionary finding was that weekly variability in home measurements of FVC, UCSD-SOBQ, and dyspnoea VAS was high. For FVC, 8% of patients demonstrated more than 10% variability in measurement from week to week, despite the high correlation of office-based and home-based measurement. This suggests that clinical confirmation of FVC decline, particularly in the absence of other findings of disease progression (e.g. worsened dyspnoea, oxygenation, or radiographic appearance) is essential to protect against the impact of random measurement error on the clinical assessment of disease course. Week-to-week effects of random measurement error may also explain the lack of correlation seen here between changes in FVC and changes in dyspnoea. Alternatively, the lack of

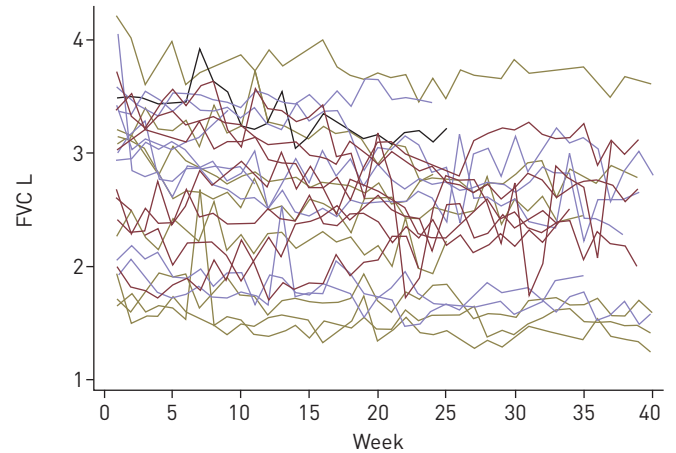


FIGURE 3 Individual weekly variability in forced vital capacity (FVC) over the study period.

correlation may be explained by the complexity of self-reported dyspnoea in patients with IPF, as dyspnoea is a symptom of multifactorial aetiology. Short-term fluctuations may be affected by factors other than FVC, and this warrants further study.

Our study has some limitations. This was an exploratory study and the sample size was small, although large enough to reveal nuanced statistical implications of repeated measures modelling. Our study cohort was established from an academic referral centre for ILD, and our findings may not be broadly generalisable. Larger and more diverse populations should be studied to further characterise the role of home monitoring in IPF. We cannot conclude whether one spirometry device performs better than another, or whether daily, weekly or some other frequency of home monitoring provides the optimal number of data. We arbitrarily selected weekly rather than daily spirometry with the goal of efficiently addressing our study hypothesis and achieving balance between data acquisition and patient burden.

In summary, we believe our findings demonstrate that weekly home assessment of FVC and dyspnoea is reliable, feasible and informative over the course of 24 weeks in patients with IPF. Home measurement of clinical endpoints in clinical trials has the potential to greatly improve analytical efficiency of clinical trials by reducing the sample size required, but this depends on the outcome-specific correlation structure of the repeated measurements – something that warrants careful study in future clinical trials through the incorporation of home-assessment outcomes as exploratory endpoints.

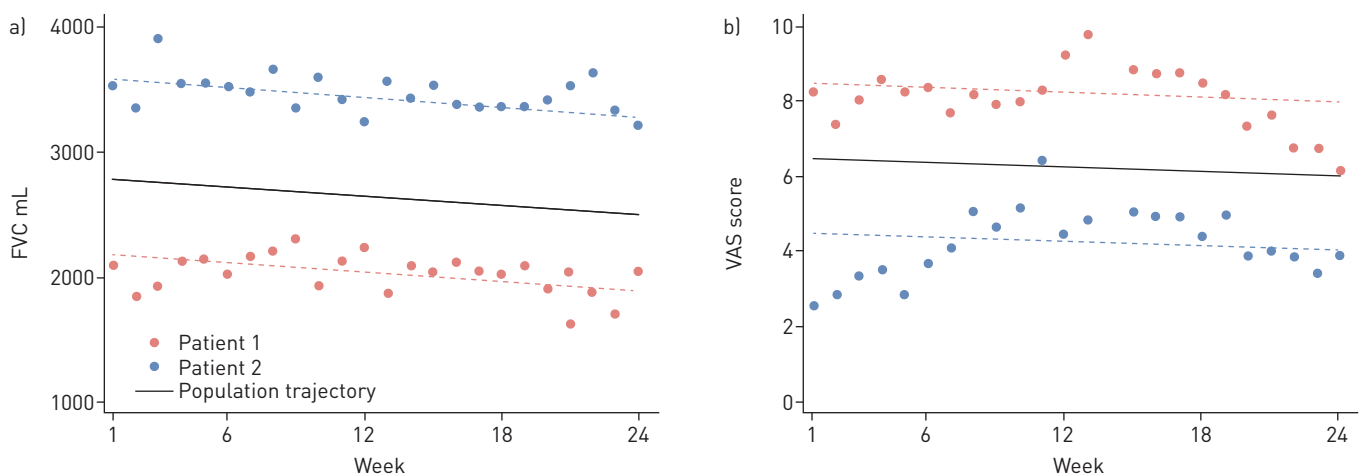


FIGURE 4 Weekly home measurements demonstrating correlation structures. a) Hand-held forced vital capacity (FVC) measures over time for two simulated patients typical of the sample. The circles represent each weekly FVC measure and the dotted lines represent the individual's modelled linear trajectory over 24 weeks. These plots demonstrate a compound symmetrical correlation structure: that is, the correlation between any two measures within an individual is independent of the time lag between the two measures. b) Home dyspnoea visual analogue scale (VAS) scores over time for two simulated patients typical of the sample. The circles represent each weekly VAS measure and the dotted lines represent the individual's modelled linear trajectory over 24 weeks. The plots demonstrate a declining correlation structure: that is, the correlation between any two measures within an individual decreases as the time between the two measures increases.

Acknowledgements

We thank the patients from the UCSF ILD programme for their time and commitment to clinical research. We also thank Rachel Stewart, Darren Leung, Maya Lalosh and Archer Eller from UCSF for their administrative support of this project.

References

- 1 Raghu G, Collard HR, Egan JJ, *et al.* An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med* 2011; 183: 788–824.
- 2 Ley B, Collard HR, King TE Jr. Clinical course and prediction of survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2011; 183: 431–440.
- 3 Martinez FJ, Safrin S, Weycker D, *et al.* The clinical course of patients with idiopathic pulmonary fibrosis. *Ann Intern Med* 2005; 142: 963–967.
- 4 Richeldi L, du Bois RM, Raghu G, *et al.* Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med* 2014; 370: 2071–2082.
- 5 King TE Jr, Bradford WZ, Castro-Bernardini S, *et al.* A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *N Engl J Med* 2014; 370: 2083–2092.
- 6 Karimi-Shah BA, Chowdhury BA. Forced vital capacity in idiopathic pulmonary fibrosis – FDA review of pirfenidone and nintedanib. *N Engl J Med* 2015; 372: 1189–1191.
- 7 Ryerson CJ, Berkeley J, Carrieri-Kohlman VL, *et al.* Depression and functional status are strongly associated with dyspnea in interstitial lung disease. *Chest* 2011; 139: 609–616.
- 8 Vittinghoff EGD, Shiboski SC, McCulloch CE. *Regression Methods in Biostatistics*. New York, Springer, 2011.
- 9 Russell AM, Adamali H, Molyneaux PL, *et al.* Daily home spirometry: an effective tool for detecting progression in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2016; 194: 989–997.
- 10 Miller MR, Hankinson J, Brusasco V, *et al.* Standardisation of spirometry. *Eur Respir J* 2005; 26: 319–338.
- 11 Eakin EG, Resnikoff PM, Prewitt LM, *et al.* Validation of a new dyspnea measure: the UCSD Shortness of Breath Questionnaire. University of California, San Diego. *Chest* 1998; 113: 619–624.
- 12 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307–310.
- 13 King TE Jr, Brown KK, Raghu G, *et al.* BUILD-3: a randomized, controlled trial of bosentan in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2011; 184: 92–99.
- 14 Noble PW, Albera C, Bradford WZ, *et al.* Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. *Lancet* 2011; 377: 1760–1769.
- 15 Demedts M, Behr J, Buhl R, *et al.* High-dose acetylcysteine in idiopathic pulmonary fibrosis. *N Engl J Med* 2005; 353: 2229–2242.
- 16 Collard HR, Bradford WZ, Cottin V, *et al.* A new era in idiopathic pulmonary fibrosis: considerations for future clinical trials. *Eur Respir J* 2015; 46: 243–249.