Jee Adelle Siew-Ying (Orcid ID: 0000-0002-2844-4370)
Youssef Peter (Orcid ID: 0000-0002-9010-3999)

TITLE

A composite serum biomarker index for the diagnosis of systemic sclerosis interstitial lung disease: a multicentre, observational, cohort study

SUB-TITLE: Composite serum biomarker index for SSc-ILD diagnosis

AUTHORS

Adelle S Jee, PhD[1,2,3]

Iain Stewart, PhD[4,5]

Peter Youssef, PhD[2,6]

Stephen Adelstein, PhD[2,7,8]

Donna Lai, PhD[9]

Sheng Hua, PhD[9]

Wendy Stevens, FRACP[10]

Susanna Proudman, FRACP[11,12]

Gene-Siew Ngian, PhD[13,14]

Ian N Glaspole, PhD[3,15,16]

Yuben P Moodley, PhD[3,17,18]

Jane F Bleasel, PhD[19]

Sacha Macansh, BAppSc[20]

Mandana Nikpour, PhD[10,21]

Joanne Sahhar*, FRACP[13,14]

Tamera J Corte*, PhD[1,2,3]

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/art.42491

On behalf of the Australian Scleroderma Cohort Study, Australian Scleroderma Interest Group, Australian Idiopathic Pulmonary Fibrosis Registry, and associated investigators[†]

*Co-senior authors

[†]*Additional investigators listed in the appendix and separate file provided of qualification, affiliations and address*

AFFILIATIONS and ADDRESSES

1. Department of Respiratory Medicine, Royal Prince Alfred Hospital, Level 11 Building 75, Missenden Road, Camperdown, Sydney NSW 2050, Australia
2. School of Medicine, University of Sydney Faculty of Medicine and Health, Sydney NSW 2006, Australia
3. NHMRC Centre of Research Excellence in Pulmonary Fibrosis, Australia
4. National Heart and Lung Institute, Imperial College, London, Guy Scadding Building, Brompton Campus, Cale Street, London, SW3 6LR, United Kingdom
5. NIHR Biomedical Research Centre, Respiratory Research Unit, University of Nottingham, Nottingham, United Kingdom
6. Institute of Rheumatology and Orthopaedics, Royal Prince Alfred Hospital, Level 4, QEII Building 59 Missenden road, Camperdown, Sydney NSW 2050, Australia
7. Department of Clinical Immunology and Allergy, Royal Prince Alfred Hospital, Level 4 Building 75, Missenden road, Camperdown, Sydney NSW 2050, Australia
8. Central Immunopathology Laboratory, NSW Health Pathology, NSW, Australia
9. Bosch Institute, Molecular Biology Facility, University of Sydney, Anderson Stuart Building F13, Sydney NSW 2006, Australia

10. Department of Rheumatology, St Vincent's Hospital Melbourne, 41 Victoria Parade, Fitzroy Victoria 3065, Australia
11. Rheumatology Unit, Royal Adelaide Hospital, Port Road, Adelaide SA 5000, Australia
12. Discipline of Medicine, University of Adelaide, Adelaide SA 5000, Australia
13. Department of Inflammatory Diseases, Monash Health, 246 Clayton Rd, Clayton 3168, Melbourne Victoria 3168, Australia
14. Department of Medicine, Monash University, Department of Rheumatology, Block E, Level 5, Monash Health, 246 Clayton Road, Clayton, Victoria 3168, Australia
15. Department of Medicine, Central Clinical School, Monash University, Commercial Road, Melbourne Victoria 3004, Australia
16. Department of Respiratory Medicine, Alfred Hospital, Commercial Road, Melbourne Victoria 3004, Australia
17. Department of Respiratory Medicine, Fiona Stanley Hospital, 11 Robin Warren Drive Murdoch, Perth WA 6150, Australia
18. Institute for Respiratory Health, University of Western Australia, QEII Medical Centre, Level 2, 6 Verdun Street, Nedlands, WA 6009 Perth, Asutralia
19. School of Medicine and Public Health, Faculty of Health Medicine and Wellbeing, The University of Newcastle, Callaghan NSW 2308, Australia
20. Lung Foundation Australia, L2 11 Finchley Street, Milton Queensland 4046, Australia
21. Department of Medicine, The University of Melbourne, St. Vincent's Hospital, Level 3, 35 Victoria Pde, Fitzroy, Melbourne Victoria 3065, Australia

FUNDING

Research grants for this study were received from Arthritis Australia and Roche Pharmaceuticals. ASIG and the ASCS is supported by Janssen (Actelion), Scleroderma

Australia, Scleroderma Victoria, Arthritis Australia, Musculoskeletal Australia, Australian Rheumatology Association, St. Vincent's Hospital Melbourne Information Technology Department and the Scleroderma Clinical Trials Consortium, and Bayer. Lung Foundation Australia facilitates the Australian IPF Registry with the generous support of unrestricted educational grants from Foundation partners Roche Products, Pty. Limited and Boehringer Ingelheim. Institutional and in-kind support was provided by the Department of Allergy and Clinical Immunology and Department of Respiratory at the Royal Prince Alfred Hospital, the University of Sydney, and the NHMRC Centre of Research Excellence in Pulmonary Fibrosis, Australia, which is funded by the NHMRC and supported by Foundation partner Boehringer Ingelheim and Program Partners Roche and Galapagos. The funders of the study had no role in the study design, data collection, analysis, interpretation or writing of the report.

DISCLOSURES

ASJ reports research grants from the Lung Foundation Australia and University of Sydney. TJC reports grants and consulting fees from Boehringer Ingelheim, Roche, BMS, Vicore, and Devpro; grants from Gilead, Bayer and Intermune; advisory board fees from Roche, BMS, Boehringer Ingelheim, Vicore, Ad Alta, Bridge Biotherapeutics, and DevPro. SP reports grants, speaker fees and advisory board fees from Actelion Australia (Janssen); grant funding from Glaxo Smith Kline; and clinical trial and advisory board fees from Boehringer-Ingelheim. PY reports institutional grants from Arthritis Australia. MN reports grants from Acetlion (Janssen) and GSK; speaking fees from Actelion (Janssen), GSK, Pfizer and UCB; consulting fees from Actelion (Janssen), Astra Zeneca, Boehringer-Ingelheim, Certa Therapeutics, Eli Lilly, GSK, Janssen and UCB; and holds an National Health and Medical Research Council of Australia Investigator Grant (GNT 1176538). JS reports advisory board fees for Boehringer Ingelhgeim. YM reports advisory board fees for Boehringer-Ingelheim and Roche. All other authors declare no competing interests.

CORRESPONDING AUTHOR:

Dr Adelle S Jee

Address: Royal Prince Alfred Hospital, Department of Respiratory Medicine, Royal Prince Alfred Hospital, Level 11 Building 75, Missenden Road, Camperdown, Sydney NSW 2050, Australia

Email: adelle.jee@health.nsw.gov.au

Telephone: +61421 697 747

ORCID ID: 0000-0002-2844-4370

WORD COUNT (manuscript body): 3401

Abstract word count: 207

KEY WORDS: Biomarkers; systemic sclerosis; pulmonary; interstitial lung disease

ABSTRACT

Objective

In patients with systemic sclerosis (SSc), we investigated composite serum biomarker panels for the diagnosis and risk-stratification of SSc-associated interstitial lung disease (SSc-ILD).

Methods

Twenty-eight biomarkers were analysed in 640 participants: 259 with SSc-ILD and 179 SSc-controls without ILD (Australian Scleroderma Cohort Study), 172 idiopathic pulmonary fibrosis (IPF)-controls (Australian IPF Registry), and 30 healthy controls. A composite index was developed from biomarkers associated with ILD in multivariable analysis derived at empirical thresholds. Performance of the index to identify ILD, and specifically SSc-ILD, and its association with lung function, radiological extent, health-related quality of life (HRQoL) were evaluated in derivation and validation cohorts. Biomarkers to distinguish SSc-ILD from IPF-controls were identified.

Results

A composite biomarker index, comprising SP-D, Ca15-3 and ICAM-1, was strongly associated with SSc-ILD diagnosis, independent of age, sex, smoking and lung function (index=3: pooled adjusted OR 12.72, 95%CI 4.59–35.21, $p<0.001$). The composite index strengthened the performance of individual biomarkers for SSc-ILD identification. In SSc patients, a higher index was associated with worse baseline disease severity (index=3 relative to index=0: adjusted absolute change in FVC% $^{-}$17.84% and DLCO% $^{-}$20.16%, both $p<0.001$).

Conclusion

A composite serum biomarker index, comprising SP-D, Ca15-3 and ICAM-1 may improve the identification and risk-stratification of ILD in SSc patients at baseline.

INTRODUCTION

Interstitial lung disease (ILD) comprises a heterogeneous group of parenchymal lung disorders characterised by varying degrees of inflammation and fibrosis. ILD is detectable in up to 80% of patients with systemic sclerosis (SSc), and a leading cause of morbidity and mortality [1]. Risk factors for SSc-ILD include older age, male sex, African American ethnicity, diffuse cutaneous SSc, short disease duration and positive anti-Scl70 autoantibodies, whilst anti-centromere autoantibodies confer a lower risk [2]. However, a definitive diagnosis of SSc-ILD may remain elusive until after irreversible lung injury has occurred. Timely, accurate identification of SSc-ILD to guide early management decisions is increasingly important, with effective therapies to stabilise and slow disease progression [3-7].

To reduce diagnostic delays, screening SSc patients with high resolution computed tomography (HRCT) and regular pulmonary function testing has been advocated [7]. The risk of ILD is highest in the first three years after SSc onset, but can develop several years later [2]. The cumulative radiation exposure from repeated HRCT screening poses an unacceptable risk in a predominantly young to middle-aged, female population. Alternative modalities, including low dose CT and lung ultrasound, require further validation [8, 9]. Lung function testing may not detect change until after significant fibrosis has become established, and extra-parenchymal factors, especially pulmonary hypertension (PH), skin thickening, and muscle weakness, may confound results.

To address these challenges, serum biomarkers have been explored as minimally invasive markers to identify SSc-ILD. Molecular biomarkers identified in SSc-ILD and its idiopathic counterpart, idiopathic pulmonary fibrosis (IPF), have provided mechanistic insights into the pathogenic pathways that drive fibrogenesis in these clinically distinct ILDs. To date, small

sample sizes, the cost and access to specialised assays, complex models and a lack of validation across diverse populations have hampered clinical translation. We hypothesised that a combination of serum biomarkers would more accurately reflect the heterogeneity of SSc-ILD and assist the identification of ILD in diverse SSc populations.

We assessed the ability of 28 prespecified blood biomarkers to identify ILD (SSc-ILD and a control group of IPF), using development and validation cohorts derived from the Australian Scleroderma Cohort Study (ASCS), Australian IPF registry (AIPFR), and healthy controls. Our primary objective was to identify a panel of serum biomarkers able to accurately distinguish SSc-ILD from non-ILD controls at baseline. Secondary objectives included evaluation of the biomarker panel for disease severity and health-related quality of life (HRQoL).

METHODS

SSc participants

The ASCS is a national, multicentre, longitudinal observational cohort established in 2007 by the Australian Scleroderma Interest Group (ASIG) to study cardiopulmonary outcomes in SSc [10]. All Australian physicians are invited to refer SSc patients to 12 national screening centres. Patients fulfilling American College of Rheumatology/European League Against Rheumatism classification for SSc are eligible for enrolment [11]. Participants provide written consent at recruitment. Utilising a standardised protocol, comprehensive de-identified data are recorded on consecutive patients at baseline and every 12 months including clinical assessment, investigations, immunomondulatory therapy, HRQoL, and outcomes (vitality, lung transplantation). Disease extent on HRCT was available on a subset of patients from prior studies. Selected centres have baseline and serial serum stored in a linked biobank. Clinical management remains as per the referring physician.

SSc participants were identified through the ASCS. Baseline was defined as the time of first serum collection. All patients with sufficient sera, HRCT confirming the ILD status, and lung function within 12 months of baseline at the time of study commencement (1 February, 2017) were included. SSc-ILD was defined as SSc with radiologically-confirmed ILD on HRCT. SSc-controls, age and sex-matched with SSc-ILD patients, included SSc patients without ILD on HRCT, and FVC ≥85%. Patients with insufficient data to confirm ILD status or serum (see "serum samples and biomarker measurement") were excluded.

Non-SSc participants

To control for non-SSc ILD, IPF-controls were identified from the AIPFR – a national, multicentre, observational registry established in 2012. IPF diagnoses were re-reviewed by a central ILD multidisciplinary meeting in accordance with American Thoracic Society/European Respiratory Society/Japanese Respiratory Society/Latin American Thoracic Association guidelines [12]. Participants provide written consent at recruitment. Comprehensive clinical and outcome data are recorded at baseline and every six months. A subset of patients have baseline and serial serum stored in linked biobanks. All AIPFR patients with sufficient baseline serum at the time of study commencement were included.

Thirty healthy controls (no history of smoking, lung disease or major comorbidities) were identified at a tertiary referral centre, consented for participation and a single serum sample taken.

The ASCS and AIPFR have human research ethics approval from all participating centres and approval for this study was granted by the Research Ethics and Governance Office (Protocol numbers X16-0311 & LNR/16/RPAH/406 RPAH zone, 99057A Monash).

Definitions

"ILD" included all patients with SSc-ILD and IPF-controls for the purposes of analysis. SSc disease duration was defined as time from first non-Raynaud symptom until baseline. All percentage predicted values for forced vital capacity (FVC%) and diffusing capacity for carbon monoxide corrected for haemoglobin (DLCOc%) were calculated from absolute measurements for FVC and DLCO, using the *Quanjer* Global Lung Initiative 2012 and *Miller* 1983 predictive equations respectively [13, 14]. Composite physiologic index (CPI), a model validated for mortality prediction in IPF and extrapolated to other ILDs, was calculated from clinical and physiological variables [15]. Disease extent was defined as per Goh *et al.*, with "limited disease" considered <20% involvement on HRCT and "extensive disease" >20% involvement or FVC <70% in indeterminate cases [16]. Immunosupressive therapy included past or current treatment with any of corticosteroids, mycophenolate, azathioprine, cyclophosphamide, methotrexate, rituximab, TNFα antagonists, tocilizumab, abatacept, or leflunomide. HRQoL in SSc patients was reported using Scleroderma Health Assessment Questionnaire (SHAQ) and 36-item short form survey instrument (SF36) [17, 18].

Follow-up data (vitality, lung transplantation) were censored on 1 May, 2019.

Serum samples and biomarker measurement

Serum samples were obtained from the ASCS and AIPFR biobanks and stored at -80°C until assayed. Sera with >2 freeze-thaw cycles were excluded. Twenty-eight biomarkers were selected by comprehensive literature review [19], and analysed at a single laboratory following

manufacturer's protocols by magnetic Luminex (SP-D, MMP-1, MMP-3, MMP-7, MMP-12, TIMP-1, Ca15-3, periostin, CCL-2, IL-6, VEGF, IL-8, CXCL-10, CXCL-12, CXCL-13, E-selectin, CXCL-4, CCL-18, ICAM-1, VCAM-1), bead (TGF-β1, -β2, -β3) and enzyme linked immunosorbent assay (ELISA; amphiregulin, KL-6, fibulin-1, LOXL-2, ET-1). Non-abbreviated biomarker labels and assay details are shown in Table S1. Any biomarkers with >25% of results below detectable range in initial analysis were re-tested with higher sensitivity assays. Thereafter, biomarkers with ≥80% missing data were excluded from analysis (n=2: KL-6, MMP-12). Individuals with missing biomarker values were imputed at half the lower limit of detection, with all such values below the lower limit.

Statistical analysis

All study participants were randomly partitioned to derivation and validation cohorts based on diagnosis to create largely balanced cohorts for development and internal replication, with the exception of healthy controls who were all considered in the derivation cohort (further detail in Supplementary Methods in the online data supplement).

Continuous biomarker measures with an area under the receiver operator characteristic (ROC) curve (AUC) >0.6 in univariable analyses were assessed for an empirical threshold in the derivation cohort using Youden index to maximise the sum of sensitivity and specificity. Biomarker thresholds with an AUC >0.6 on ROC analyses were included for further multivariable analysis. Least Angled Regression (LARS) was used to select a parsimonious number of biomarkers that could predict ILD outcome respective to a-priori defined clinical covariates (baseline age, sex, smoking, FVC%) to retain in a multivariable model. The optimum prediction error for balance between bias and variance in LARS was determined using the Cp statistic, biomarkers beyond the optimal error were not retained. Association of the retained

biomarker threshold with ILD was tested univariably, and in mutually adjusted models with retained biomarkers and clinical covariates. A biomarker index was developed comprising of retained biomarkers significantly associated with ILD in the multivariable model, ranging from 0 (no biomarkers above threshold) to 3 (three biomarkers above threshold).

Univariable and multivariable logistic regression (adjusted for baseline age, sex, smoking, FVC%) were used to evaluate associations between the index and ILD. In a sensitivity analysis restricted to SSc diagnoses, multivariable adjustment was further extended to include disease duration and Scl-70 positivity. Generalised linear models were constructed to investigate associations of the index with physiology, disease extent and HRQoL metrics in univariable and multivariable analyses adjusted for clinical covariates and immunosuppression. Analyses were performed separately in the derivation blinded to the validation cohort, with multilevel modelling and random intercept over cohort to test associations in pooled analyses.

Continuous variables are reported as mean (SD, standard deviation) or median (IQR, interquartile range) as appropriate, and categorical variables as an absolute number (relative frequency). Regression models are presented with 95% confidence intervals (95%CI). P-values <0.05 were considered statistically significant. All statistical analysis was performed using Stata statistical software (versions 14.2–16.1).

RESULTS

Baseline characteristics

A total 640 patients met inclusion criteria, including 259 with SSc-ILD (40.5%), 179 SSc-controls (27.9%), 172 IPF-controls (26.9%) and 30 healthy controls (4.7%). Overall mean age was 60 years, 32.7% male, 55.9% never smokers with mild-to-moderate lung function

impairment (mean FVC 87.9% and DLCO 59.2% predicted). Lung function impairment was less severe in SSc-ILD patients compared with IPF-controls, but more severe compared to SSc-controls (Table 1). Disease duration did not differ between SSc-ILD and SSc-controls (p=0.765). SSc-ILD patients compared with SSc-controls demonstrated a higher prevalence of anti-Scl70 autoantibodies (38.5% vs. 9.3% respectively, p<0.001), lower prevalence of anti-centromere autoantibodies (18.8% vs. 49.4% respectively, p<0.001), and were more likely to have received immunomodulatory therapy (63.7% vs. 46.9% respectively, p=0.001; see Table 1).

The derivation cohort included 292 patients (111 SSc-ILD [38.0%], 77 SSc-controls [26.4%], 74 IPF-controls [25.3%], 30 healthy controls [10.3%]) and the validation cohort included 348 patients (148 SSc-ILD [42.5%], 102 SSc-controls [29.3%], 98 IPF-controls [28.2%]). Derivation and validation cohorts demonstrated no difference in baseline characteristics or lung function (Table S2).

Outcomes

Median follow-up time was 60.7 months (IQR 53.2–73.3) and 69.7 months (IQR 56.8–74.9) in the derivation and validation cohorts respectively. There were 166 deaths (25.9% overall; 70 SSc-ILD, 21 SSc-controls, 75 IPF-controls) and 25 lung transplants (3.9% overall; 3 SSc-ILD, 22 IPF-controls).

Development and validation of the biomarker index for ILD and SSc-ILD

Median biomarker concentrations by patient cohort are shown in Table S3. Development of the biomarker index is summarised in Figure 1. ROC curve analysis, empirical biomarker

thresholds and LARS model specification for the selection of biomarkers to include in the biomarker index are shown in Table S4 and Figure S1.

In univariable logistic regression, SP-D, Ca15-3, ICAM-1, TIMP-1 and MMP-7 were associated with ILD (Table 2). After multivariable adjustment, SP-D, Ca15-3 and ICAM-1 remained associated with ILD for inclusion in the final index. A biomarker index score (0–3) was calculated from the number of these biomarkers above the empirical threshold in any combination, and taken forward for validation. Distribution of index scores by patient group is shown in Figure 2.

Performance of Biomarker Index to distinguish ILD and SSc-ILD

Performance of the biomarker index to discriminate ILD (SSc-ILD or IPF-controls) from no-ILD (SSc without-ILD and healthy controls) is shown in Table 3. After adjustment for age, sex, smoking and FVC%, an index of ≥2 was strongly associated with ILD in derivation, validation and pooled analyses (index=2 pooled adjusted OR [aOR] 3.59, 95%CI 1.91–6.77, p<0.001; index=3 aOR 15.61, 95%CI 5.81–41.61, p<0.001; unadjusted OR shown in Table S5).

In sensitivity analysis limited to SSc patients, the biomarker index remained robust to effectively discriminate SSc-ILD from SSc-controls. An index of 3 was strongly associated with SSc-ILD in derivation, validation and pooled analyses (aOR 12.72 in pooled analysis, 95%CI 4.59–35.21, p<0.001; Table 3). An index of 2 was also strongly associated with SSc-ILD in the derivation cohort and pooled data (aOR 3.24, 95%CI 1.67–6.30, p=0.001). Similar results and direction of effect were observed in multivariable logistic regression further adjusted for SSc disease duration and Scl-70 positivity (Table S6) and when stratified by diffuse-cutaneous and limited-cutaneous SSc (Table S7).

Biomarker index and disease severity in SSc

In SSc patients, a biomarker index ≥2 was associated with more severe physiological impairment by FVC%, DLCOc% and CPI (Figure 3; Table S8). In pooled analysis adjusted for age, sex and smoking, an index of 3 was associated with a lower FVC% (⁻17.84%, 95%CI ⁻24.73–⁻10.94, p<0.001), DLCOc% (⁻20.16%, 95%CI ⁻26.23–⁻14.08, p<0.001) and higher CPI (17.77 points, 95%CI 12.85–22.69, p<0.001) compared with an index of 0. The ability of the index to distinguish lung function was not impacted by immunomodulatory therapy (Table S9).

Of 139 SSc-ILD patients with available HRCT-extent data, extensive disease was present in 48 patients (34.5%; derivation n=18 [30%], validation n=30 [38.0%]). An index ≥2 was associated with extensive disease in pooled analysis (aOR 1.33–1.46; Table S10).

The biomarker index was not associated with HRQoL metrics, SHAQ or SF36 (Table S11).

DISCUSSION

In 640 patients with well-defined SSc-ILD, SSc-without ILD, IPF and healthy controls, we identified and validated a simple three-biomarker index, comprised of SP-D, ICAM-1 and Ca15-3, as a baseline marker of ILD and disease severity, independent of clinical variables. To our knowledge, our study is the first and largest multicentre study to evaluate a comprehensive array of serum biomarkers in SSc-ILD, including development and validation analysis. Our findings support the use of a composite serum biomarker index as a minimally invasive, objective tool to support current diagnostic modalities and multidisciplinary discussion for ILD diagnosis and risk-stratification in SSc.

We identified three independent serum biomarkers associated with ILD, (SP-D, Ca15-3, ICAM-1), which as a composite index were able to robustly discriminate SSc-ILD from SSc-controls at baseline, independent of age, sex, smoking and FVC%. SP-D plays an important role in innate immune defence, Ca15-3 in epithelial cell recognition and adhesion, and ICAM-1 mediates migration of transendothelial leucocytes into the interstitium and activation of downstream signalling pathways including TGF-β1 [20-22]. Our findings are consistent with prior studies identifying elevated levels of these biomarkers in response to epithelial and endothelial cell injury, identified as a common trigger of fibrogenesis in both SSc-ILD and IPF [23]. Similar to the PROFILE study in IPF patients, we found that SP-D and epithelial-based biomarkers were superior to markers of extracellular matrix remodelling, including MMP-7, to distinguish ILD from controls [24]. Importantly, building on prior studies of SP-D, Ca15-3 and ICAM-1 as individual metrics for SSc-ILD diagnosis, we demonstrated that combining these biomarkers into a simple composite index strengthened their performance characteristics in a robust, easy to calculate tool [19]. We believe this composite approach is more reflective of the complex clinical and molecular heterogeneity of SSc-ILD, and less liable to environmental or genetic confounders compared with individual markers. This may improve the accuracy and generalisability of the index in heterogeneous SSc populations, a concept supported by White et al.'s biomarker index for IPF [25]. Assessment in external cohorts and combination in a multidimensional model with existing demographic, physiologic, serologic and histo-pathologic risk factors for SSc-ILD is vital.

The inclusion of Ca15-3 in our biomarker index is of particular interest. Ca15-3, most commonly recognised as a tumour marker, is an epitope of mucin 1 (MUC-1), expressed on the surface of various epithelial cells including type II alveolar, breast, hepatic, gastrointestinal and genitourinary cells [26]. Ca15-3 is encoded for by the same MUC-1 gene as KL-6, which is

among the most frequently reported biomarkers in ILD [26]. Despite promise, several limitations have limited wider clinical application of KL-6. Specialised assays for its measurement are not widely available outside of the research setting in many countries, and there is a lack standardisation or validation in geographically diverse cohorts, with markedly different KL-6 levels reported in ethnically and genotypically disparate populations [27]. Supporting our findings, prior studies have demonstrated elevated Ca15-3 in SSc-ILD and other ILDs, with good sensitivity and specificity for differentiating ILD from no-ILD, and an association with ILD-extent by HRCT and FVC [22, 28-31]. Good correlation between Ca15-3 and KL-6 levels in ILD patients has been described and robust validation in broad SSc populations is required [26]. The expression of Ca15-3 in non-pulmonary and malignant disease may limit its specificity alone, providing further rationale for its combination with other biomarkers in a combined index. In sum, our data supports further investigation of Ca15-3 as a lower cost, more accessible surrogate for KL-6 in future biomarker panels for SSc-ILD assessment.

In our study, higher biomarker index scores were strongly associated with more severe lung function impairment, increased disease extent on HRCT (acknowledging small numbers), and higher CPI, independent of clinical variables. The positive relationship between the number of elevated biomarkers and magnitude of ILD risk may provide an additional benefit over traditional risk factors, such as Scl-70, which denotes an increased risk of SSc-ILD, but not the degree. Baseline SSc-ILD severity assessment is critical to early risk-stratification, prognosis and clinical management. Extensive disease on HRCT and lower FVC at baseline are risk factors for progressive SSc-ILD and increased mortality [1, 2, 7, 16]. However, visual HRCT evaluation is subject to inter-observer variability, and automated quantitative methods await validation [32]. Lung function in SSc-ILD patients may be confounded by concomitant PH

contributing to a reduced DLCO, or muscular weakness and/or thoracic skin thickening reducing lung volumes independent of parenchymal involvement. Serum biomarkers are unlikely to replace HRCT or lung function for diagnosis or severity assessment in SSc-ILD. Rather, the biomarker index may provide an additional objective, point-of-care test to guide ILD assessment in SSc patients when standard modalities are indeterminate or confounding factors present. The additional benefit of biomarkers further to HRCT and lung function for ILD diagnosis and risk stratification in multidimensional models requires external validation in prospective SSc cohorts.

The need to identify early, progressive SSc-ILD has been amplified by the emergence of antifibrotics as effective disease-modifying agents in non-IPF progressive-fibrosing ILDs [3-6]. Analysis of our biomarker panel to predict progressive disease is ongoing. The utility of serial biomarker index measurements to monitor disease severity over time requires longitudinal assessment. However, this should not diminish the need to accurately identify and treat the primary clinical diagnosis at baseline. Immunomodulation remains first-line therapy in SSc-ILD, with antifibrotic therapy for progressive-ILD in non-IPF ILD based on disease behaviour after conventional therapy. Furthermore, antifibrotics for non-IPF ILDs are not universally available. The utility of biomarkers to guide the selection of SSc-ILD patients for immunomodulatory therapy, antifibrotics, or both, remains an urgent unmet need.

Our study has several limitations. Our two large cohorts were recruited from observational registry data, thus diagnostic procedures were performed when clinically indicated and datasets had variable timing between sera collection, lung function and HRCT testing. Radiological subtypes of SSc-ILD (e.g. usual interstitial pneumonia (UIP) pattern, non-specific interstitial pneumonia (NSIP), etc.) could not be accounted for. Nevertheless, sensitivity analysis of the

biomarker index limited to SSc patients and excluding IPF-controls, the latter all with radiological UIP by definition, remained robust for SSc-ILD diagnosis. This suggests that a UIP pattern was less likely to be driving these findings. The precise impact of immunomodulation on biomarker levels in this cohort could not be determined. Considerable evolution of SSc and ILD treatments over time, a lack of universal guidelines for the treatment of ILD in SSc, and considerable variability between centres precluded analysis. Nevertheless, immunomodulation did not have an impact on the biomarker index's ability to distinguish lung function, suggesting therapy was not driving biomarker effects. Studies suggest most degradation of serum biomarker levels occurs after three freeze-thaw cycles [33], thus biobanked sera with >2 cycles were excluded.

Strengths of our study include the utilisation of large, well-characterised cohorts of SSc and IPF patients referred from multiple tertiary and community centres across Australia, allowing derivation of empirical thresholds and internal validation of a wide spectrum of serum biomarkers in diverse, real-world ILD populations. To improve comparability in lung function, all FVC% and DLCOc% predictive values were calculated from absolute measurements using standardised reference equations.

In conclusion, we identified a simple composite biomarker index, comprised of SP-D, Ca15-3 and ICAM-1, which strengthened the performance of individual biomarkers to robustly discriminate ILD in SSc and was associated with disease severity, independent of clinical variables. Our data supports further investigation of composite biomarker indices as objective, minimally-invasive, point-of-care tests to complement current diagnostic modalities to enhance diagnostic precision and risk-stratification of ILD in SSc. Ca15-3 may be a lower cost, more accessible surrogate for KL-6 in future biomarker assessment.

Acknowledgements

The authors would also like to acknowledge ASCS Project Officers Dr Michelle Wilson and Candice Rabusa, and all the participants and physicians who generously contribute to the ASCS and AIPFR.

References

1. Steen VD, Medsger TA. Changes in causes of death in systemic sclerosis, 1972-2002. Annals of the rheumatic diseases. 2007;66(7):940-4.

2. Nihtyanova SI, Schreiber BE, Ong VH, Rosenberg D, Moinzadeh P, Coghlan JG, et al. Prediction of pulmonary complications and long-term survival in systemic sclerosis. Arthritis rheumatol. 2014;66(6):1625-35.

3. Distler O, Highland KB, Gahlemann M, Azuma A, Fischer A, Mayes MD, et al. Nintedanib for Systemic Sclerosis–Associated Interstitial Lung Disease. New England Journal of Medicine. 2019.

4. Flaherty KR, Wells AU, Cottin V, Devaraj A, Walsh SLF, Inoue Y, et al. Nintedanib in Progressive Fibrosing Interstitial Lung Diseases. The New England journal of medicine. 2019.

5. Maher TM, Corte TJ, Fischer A, Kreuter M, Lederer DJ, Molina-Molina M, et al. Pirfenidone in patients with unclassifiable progressive fibrosing interstitial lung disease: a double-blind, randomised, placebo-controlled, phase 2 trial. The Lancet Respiratory medicine. 2019.

6. Behr J, Prasse A, Kreuter M, Johow J, Rabe KF, Bonella F, et al. Pirfenidone in patients with progressive fibrotic interstitial lung diseases other than idiopathic pulmonary fibrosis (RELIEF): a double-blind, randomised, placebo-controlled, phase 2b trial. The Lancet Respiratory Medicine. 2021;9(5):476-86.

7. Hoffmann-Vold A-M, Maher TM, Philpot EE, Ashrafzadeh A, Barake R, Barsotti S, et al. The identification and management of interstitial lung disease in systemic sclerosis: evidence-based European consensus statements. The Lancet Rheumatology. 2020;2(2):e71-e83.

8. Gasperini ML, Gigante A, Iacolare A, Pellicano C, Lucci S, Rosato E. The predictive role of lung ultrasound in progression of scleroderma interstitial lung disease. Clinical rheumatology. 2020;39(1):119-23.

9. Frauenfelder T, Winklehner A, Nguyen TD, Dobrota R, Baumueller S, Maurer B, et al. Screening for interstitial lung disease in systemic sclerosis: performance of high-resolution CT with limited number of slices: a prospective study. Annals of the rheumatic diseases. 2014;73(12):2069-73.

10. Moore OA, Goh N, Corte T, Rouse H, Hennessy O, Thakkar V, et al. Extent of disease on high-resolution computed tomography lung is a predictor of decline and mortality in systemic sclerosis-related interstitial lung disease. Rheumatology (Oxford). 2013;52(1):155-60.

11. van den Hoogen F, Khanna D, Fransen J, Johnson SR, Baron M, Tyndall A, et al. 2013 classification criteria for systemic sclerosis: an American college of rheumatology/European league against rheumatism collaborative initiative. Annals of the rheumatic diseases. 2013;72(11):1747-55.

12. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, 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(6):788-824.

13. Miller A, Thornton JC, Warshaw R, Anderson H, Teirstein AS, Selikoff IJ. Single breath diffusing capacity in a representative sample of the population of Michigan, a large industrial state. Predicted values, lower limits of normal, and frequencies of abnormality by smoking history. Am Rev Respir Dis. 1983;127(3):270-7.

14. Quanjer PH, Stanojevic S, Cole TJ, Baur X, Hall GL, Culver BH, et al. Multi-ethnic reference values for spirometry for the 3–95-yr age range: the global lung function 2012 equations. European Respiratory Journal. 2012;40(6):1324-43.

15. Wells AU, Desai SR, Rubens MB, Goh NS, Cramer D, Nicholson AG, et al. Idiopathic pulmonary fibrosis: a composite physiologic index derived from disease extent observed by computed tomography. Am J Respir Crit Care Med. 2003;167(7):962-9.

16. Goh NS, Desai SR, Veeraraghavan S, Hansell DM, Copley SJ, Maher TM, et al. Interstitial lung disease in systemic sclerosis: a simple staging system. Am J Respir Crit Care Med. 2008;177(11):1248-54.

17. Steen VD, Medsger Jr. TA. The value of the health assessment questionnaire and special patient-generated scales to demonstrate change in systemic sclerosis patients over time. Arthritis & Rheumatism. 1997;40(11):1984-91.

18. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care. 1992;30(6):473-83.

19. Jee AS, Sahhar J, Youssef P, Bleasel J, Adelstein S, Nguyen M, et al. Review: Serum biomarkers in idiopathic pulmonary fibrosis and systemic sclerosis associated interstitial lung disease – frontiers and horizons. Pharmacology & Therapeutics. 2019;202:40-52.

20. Pastva AM, Wright JR, Williams KL. Immunomodulatory roles of surfactant proteins A and D: implications in lung disease. Proc Am Thorac Soc. 2007;4(3):252-7.

21. Tsoutsou PG, Gourgoulianis KI, Petinaki E, Mpaka M, Efremidou S, Maniatis A, et al. ICAM-1, ICAM-2 and ICAM-3 in the sera of patients with idiopathic pulmonary fibrosis. Inflammation. 2004;28(6):359-64.

22. Ricci A, Mariotta S, Bronzetti E, Bruno P, Vismara L, De Dominicis C, et al. Serum CA 15-3 is increased in pulmonary fibrosis. Sarcoidosis, vasculitis, and diffuse lung diseases. 2009;26(1):54-63.

23. Herzog EL, Mathur A, Tager AM, Feghali-Bostwick C, Schneider F, Varga J. Review: Interstitial Lung Disease Associated With Systemic Sclerosis and Idiopathic Pulmonary Fibrosis: How Similar and Distinct? Arthritis rheumatol. 2014;66(8):1967-78.

24. Maher TM, Oballa E, Simpson JK, Porte J, Habgood A, Fahy WA, et al. An epithelial biomarker signature for idiopathic pulmonary fibrosis: an analysis from the multicentre PROFILE cohort study. The Lancet Respiratory Medicine. 2017;5(12):946-55.

25. White ES, Xia M, Murray S, Dyal R, Flaherty CM, Flaherty KR, et al. Plasma Surfactant Protein-D, Matrix Metalloproteinase-7, and Osteopontin Index Distinguishes Idiopathic Pulmonary Fibrosis from Other Idiopathic Interstitial Pneumonias. American Journal of Respiratory and Critical Care Medicine. 2016;194(10):1242-51.

26. Okada M, Suzuki K, Nakanishi T, Nakashima M. Serum levels of KL-6 are positively correlated with those of CA15-3 in patients with interstitial pneumonia associated with collagen diseases. Respirology (Carlton, Vic). 2006;11(4):509-10.

27. Horimasu Y, Hattori N, Ishikawa N, Kawase S, Tanaka S, Yoshioka K, et al. Different MUC1 gene polymorphisms in German and Japanese ethnicities affect serum KL-6 levels. Respiratory medicine. 2012;106(12):1756-64.

28. Celeste S, Santaniello A, Caronni M, Franchi J, Severino A, Scorza R, et al. Carbohydrate antigen 15.3 as a serum biomarker of interstitial lung disease in systemic sclerosis patients. Eur. 2013;24(7):671-6.

29. Rusanov V, Kramer MR, Raviv Y, Medalion B, Guber A, Shitrit D. The significance of elevated tumor markers among patients with idiopathic pulmonary fibrosis before and after lung transplantation. Chest. 2012;141(4):1047-54.

30. Wong RC, Brown S, Clarke BE, Klingberg S, Zimmerman PV. Transient elevation of the tumor markers CA 15-3 and CASA as markers of interstitial lung disease rather than underlying malignancy in dermatomyositis sine myositis. Journal of clinical rheumatology : practical reports on rheumatic & musculoskeletal diseases. 2002;8(4):204-7.

31. De Luca G, Bosello SL, Berardi G, Rucco M, Canestrari G, Correra M, et al. Tumour-associated antigens in systemic sclerosis patients with interstitial lung disease: association with lung involvement and cancer risk. Rheumatology. 2015;54(11):1991-9.

32. Jacob J, Bartholmai BJ, Rajagopalan S, Brun AL, Egashira R, Karwoski R, et al. Evaluation of computer-based computer tomography stratification against outcome models in connective tissue disease-related interstitial lung disease: a patient outcome study. BMC Medicine. 2016;14(1).

33. Huang W-Y, Kemp TJ, Pfeiffer RM, Pinto LA, Hildesheim A, Purdue MP. Impact of freeze-thaw cycles on circulating inflammation marker measurements. Cytokine. 2017;95:113-7.

FIGURE LEGENDS

Figure 1. Selection of biomarkers and development of the fibrosis index

ELISA = enzyme linked immunosorbent assay. ROC = receiver operator characteristic. AUC = area under the curve. LARS = least angled regression. Non-abbreviated biomarker labels and assay details are shown in Table S1.

Figure 2. Distribution of index by patient group in the derivation cohort

SSc-controls = systemic sclerosis without ILD. SSc-ILD = systemic sclerosis with ILD. IPF = idiopathic pulmonary fibrosis.

Figure 3. Association of the biomarker index with physiology

Absolute change in physiology relative to an index of 0 (red horizontal line) in derivation (blue vertical bars), validation (red vertical bars), and pooled (green vertical bars) analyses, adjusted for age, sex and smoking history. (A) Percentage predicted FVC, (B) Percentage predicted DLCO, (C) Composite physiologic index (CPI). Dots demonstrate coefficient values, vertical bars demonstrate 95% confidence intervals. FVC% = percentage predicted forced vital capacity; DLCO% = percentage predicted diffusing capacity for carbon monoxide CPI = composite physiologic index.

TABLES

Table 1. Baseline characteristics by patient group

	SSc-ILD n=259	SSc-controls n=179	IPF-controls n=172	Healthy controls n=30	TOTAL n=640	p-value* SSc-ILD vs SSc-controls	p-value* SSc-ILD vs IPF-controls
Frequency							
Derivation n(%)	111 (38.0)	77 (26.4)	74 (25.3)	30 (10.3)	292	—	—
Validation n(%)	148 (42.5)	102 (29.3)	98 (28.2)	0 (0)	348	—	—
Baseline characteristics							
Age years	57.3 (12.7)	57.9 (12.3)	70.3 (8.1)	36.8 (9.8)	60.0 (13.7)	0.628	<0.001
Sex							
Male n(%)	53 (20.5)	31 (17.3)	120 (69.8)	5 (16.7)	209 (32.7)	0.411	<0.001
Female n(%)	206 (79.5)	148 (82.7)	52 (30.2)	25 (83.3)	631 (67.3)	0.411	<0.001
Never smoker, n(%)	111 (42.9)	103 (57.5)	114 (66.3)	30 (100)	358 (55.9)	0.001	<0.001
White, n(%)	220 (84.9)	168 (93.9)	151 (87.8)	18 (60.0)	557 (87.0)	0.004	0.403
BMI kg/m^2	26.0 (5.3)	27.1 (7.0)	28.8 (4.8)	NA	27.1 (5.8)	0.105	<0.001
FVC%	86.8 (21.4)	100.8 (21.0)	77.9 (19.7)	NA	87.9 (22.5)	<0.001	<0.001
DLCOc%	59.5 (18.7)	69.5 (18.3)	48.3 (15.7)	NA	59.2 (19.4)	<0.001	<0.001
CPI	34.5 (15.3)	24.1 (14.5)	45.7 (12.4)	NA	34.6 (16.4)	<0.001	<0.001
PH n/N (%)	35/257 (13.6)	23/179 (12.9)	28/160 (17.5)	NA	—	0.886	0.325
SSc specific							
Disease duration† (years)	10.9‡ (10.2)	11.2§ (10.2)	NA	NA	—	0.765	—
Diffuse SSc, n(%)	103 (39.8)	42 (23.5)	NA	NA	—	<0.001	—
ANA positive n/N (%)	248/258 (59.8)	167/179 (40.2)	NA	NA	—	0.190	—

Anti-centromere autoantibody n/N (%)	48/255 (18.8)	88/178 (49.4)	NA	NA	—	<0.001	—
Anti-Scl-70 autoantibody n/N (%)	99/257 (38.5)	16/173 (9.3)	NA	NA	—	<0.001	—
Immunosuppression ever n(%)	165 (63.7)	84 (46.9)	NA	NA	—	0.001	—

Shown as mean(SD) unless stated. *Bonferroni adjusted p-value <0.0028 considered significant. †Time from first non-Raynaud symptom to baseline. ‡N=248. §N=171. SSc=systemic sclerosis. IPF=idiopathic pulmonary fibrosis. HC=healthy controls. BMI=body mass index. FVC%=percentage predicted forced vital capacity. DLCOc%=percentage predicted diffusing capacity for carbon monoxide corrected for haemoglobin. CPI=composite physiologic index. PH=pulmonary hypertension. NA=not applicable.

Table 2. Association of serum biomarkers for the discrimination of interstitial lung disease (SSc-ILD or IPF) at empirical thresholds

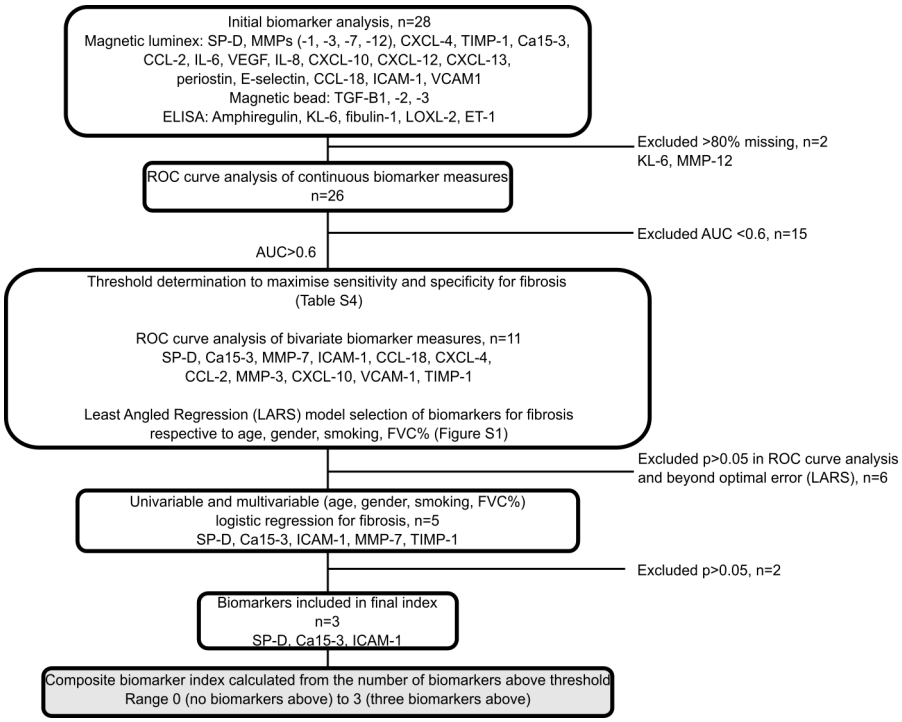
		Unadjusted			Adjusted*		
Biomarker	Threshold (pg/ml)	OR	95%CI	p	OR	95%CI	p
SP-D	32603.10	11.08	6.28–19.55	<0.001	4.21	1.99–8.92	<0.001
Ca15-3	35.90	8.11	4.71–13.96	<0.001	4.53	2.11–9.71	<0.001
ICAM-1	649531.15	4.78	2.81–8.14	<0.001	2.38	1.11–5.12	0.026
TIMP-1	149065.16	2.79	1.71–4.56	<0.001	1.97	0.90–4.29	0.088
MMP-7	1749.35	5.89	3.49–9.34	<0.001	1.22	0.55–2.71	0.621

*Mutually adjusted for biomarker threshold, age, sex, smoking, FVC%. OR=odds ratio.

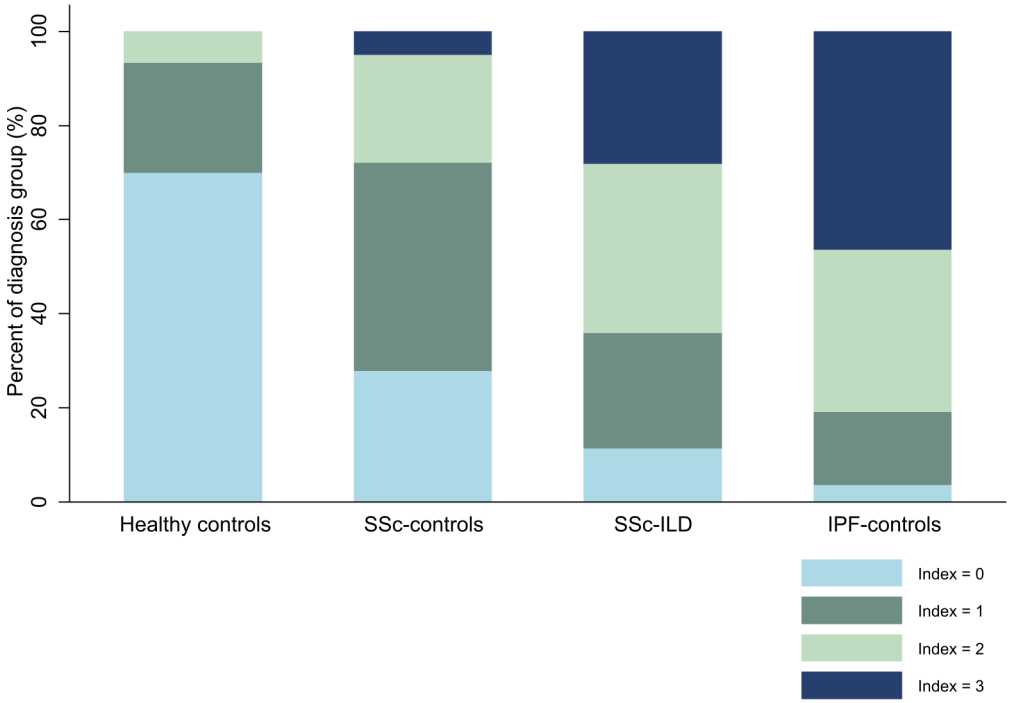
Table 3. Multivariable logistic regression of the index for a diagnosis of interstitial lung disease and SSc-ILD: derivation, validation and pooled cohorts

	Derivation			Validation			Pooled		
Index	aOR*	95%CI	p	aOR*	95%CI	p	aOR*	95%CI	p
For identification of interstitial lung disease†									
0	1	—	—	1	—	—	1	—	—
1	1.54	0.55–4.34	0.414	1.65	0.76–3.61	0.205	1.41	0.77–2.59	0.264
2	7.79	2.55–23.80	<0.001	2.40	1.11–5.18	0.025	3.59	1.91–6.77	<0.001
3	95.50	10.74–849.18	<0.001	6.72	2.01–22.47	0.002	15.61	5.85–41.61	<0.001
For identification of SSc-ILD from SSc-controls									
0	1	—	—	1	—	—	1	—	—
1	1.19	0.42–3.36	0.774	1.45	0.64–3.27	0.374	1.25	0.66–2.36	0.489
2	6.07	1.99–18.53*	0.002	1.77	0.95–5.00	0.067	3.24	1.67–6.30	0.001
3	66.33	7.45–590.96*	<0.001	5.32	1.33–21.25	0.036	12.72	4.59–35.21	<0.001

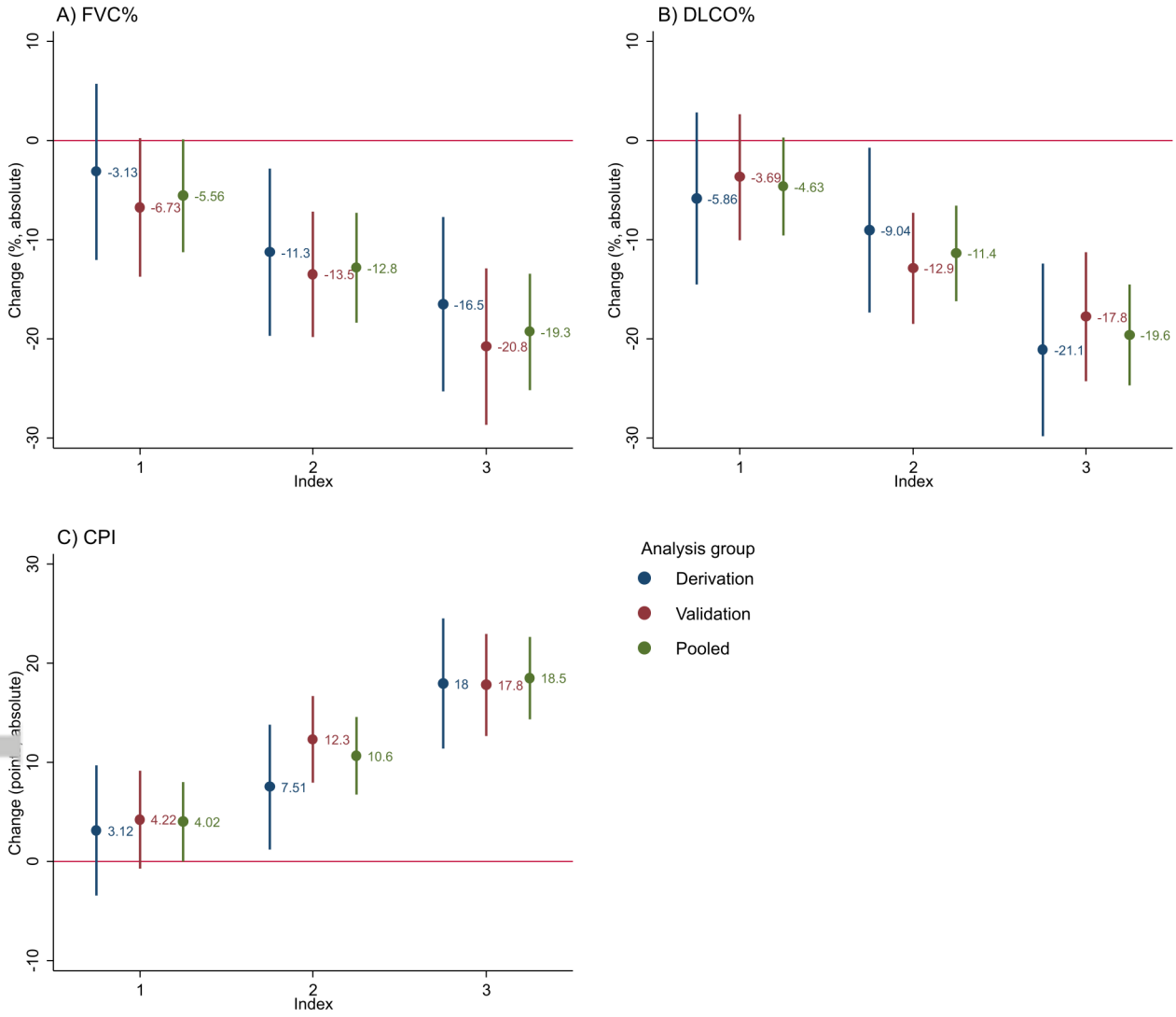
*Adjusted odds ratio (aOR) for age, sex, smoking history and FVC%. †Interstitial lung disease includes the presence of SSc-ILD or IPF; excludes healthy controls due to missing lung function. aOR = adjusted OR.



ART_42491_Figure 1. Biomarker selection.tiff



ART_42491_Figure 2. Index bar graph.tiff



ART_42491_Figure 3. Physiology plot.tiff