




Results of the standard set for pulmonary sarcoidosis: feasibility and multicentre outcomes

Nynke A. Kampstra ^{1,2,3}, Paul B. van der Nat^{1,2}, Lea M. Dijkman¹, Frouke T. van Beek³, Daniel A. Culver⁴, Robert P. Baughman⁵, Elisabetta A. Renzoni⁶, Wim Wuyts ⁷, Vasilis Kouranos⁶, Pieter Zanen⁸, Marlies S. Wijsenbeek⁹, Marinus J.C. Eijkemans¹⁰, Douwe H. Biesma^{1,11}, Philip J. van der Wees² and Jan C. Grutters^{3,8}

ABSTRACT Our study presents findings on a previously developed standard set of clinical outcome data for pulmonary sarcoidosis patients. We aimed to assess whether changes in outcome varied between the different centres and to evaluate the feasibility of collecting the standard set retrospectively.

This retrospective observational comparative benchmark study included six interstitial lung disease expert centres based in the Netherlands, Belgium, the UK and the USA. The standard set of outcome measures included 1) mortality, 2) changes in pulmonary function (forced vital capacity (FVC), forced expiratory volume in 1 s, diffusing capacity of the lung for carbon monoxide), 3) soluble interleukin-2 receptor (sIL-2R) change, 4) weight changes, 5) quality-of-life (QoL) measures, 6) osteoporosis and 7) clinical outcome status (COS). Data collection was considered feasible if the data were collected in $\geq 80\%$ of all patients.

509 patients were included in the retrospective cohort. In total six patients died, with a mean survival of 38 ± 23.4 months after the diagnosis. Centres varied in mean baseline FVC, ranging from 110 (95% CI 92–124)% predicted to 99 (95% CI 97–123)% pred. Mean baseline body mass index (BMI) of patients in the different centres varied between 27 (95% CI 23.6–29.4) $\text{kg}\cdot\text{m}^{-2}$ and 31.8 (95% CI 28.1–35.6) $\text{kg}\cdot\text{m}^{-2}$. 310 (60.9%) patients were still on systemic therapy 2 years after the diagnosis. It was feasible to measure mortality, changes in pulmonary function, weight changes and COS. It is not (yet) feasible to retrospectively collect sIL-2R, osteoporosis and QoL data internationally.

This study shows that data collection for the standard set of outcome measures for pulmonary sarcoidosis was feasible for four out of seven outcome measures. Trends in pulmonary function and BMI were similar for different hospitals when comparing different practices.



@ERSpublications

Clinical outcome data have been used to compare outcomes in pulmonary sarcoidosis patients and improve care delivery. Data collection for the standard set of outcome measures for pulmonary sarcoidosis was feasible for four out of seven outcome measures. <http://bit.ly/2F8bQ6s>

Cite this article as: Kampstra NA, van der Nat PB, Dijkman LM, *et al.* Results of the standard set for pulmonary sarcoidosis: feasibility and multicentre outcomes. *ERJ Open Res* 2019; 5: 00094-2019 [<https://doi.org/10.1183/23120541.00094-2019>].



Introduction

Reporting outcomes improves healthcare transparency and is increasingly important in daily clinical practice [1–3]. Reporting of outcomes enables quality improvement, optimises patient safety and improves care and cost-effectiveness. It is important to select outcomes consistently and apply transparent methodologies to ensure data comparison between institutions [1, 4]. The patient primarily benefits from increasing the value of care per dollar spent as the key stakeholder [2, 3]. In addition, insurers and the healthcare systems as a whole benefit from this.

Globally, the concept of value-based healthcare (VBHC) is embraced by hospitals. VBHC aims to achieve the best possible health outcomes for the lowest cost [2, 3]. This concept changes the focus from “volume” to “value” and results in tracking and comparing outcomes [5, 6]. The VBHC approach identifies practice variations and best practices in the field, which can be seen as a starting point for further quality improvement.

In sarcoidosis there is a need to define appropriate outcomes for randomised clinical trials (RCTs) and for daily clinical practice to improve the quality of care [7]. We developed a standard clinical outcome set (hereinafter referred to as the standard set) that provides insight in quality of care [8]. However, it is not known whether this set is implementable and unambiguously measurable in centres from different countries. To investigate whether this set can be used to identify clinical differences between interstitial lung disease (ILD) centres, we present retrospectively collected clinical data from six ILD expert clinics in the Netherlands, Belgium, the UK and the USA. This study was designed to assess outcomes of centres with differing patient populations and approaches to management.

The primary objective of this study was to present the standard set of clinical data across the six centres. The secondary objective was to evaluate the feasibility of retrospective data collection of the standard set in practice. Finally, the study aimed to assess whether changes in outcome at 2 years varied between the different centres.

Materials and methods

Study structure and design

This study is a multicentre retrospective observational cohort study. Six ILD expert centres participated: two from the Netherlands (St Antonius Hospital and Erasmus University Medical Center), one in Belgium (Leuven University Hospital), the UK (Royal Brompton Hospital) and two in the USA (Cleveland Clinic and Cincinnati Medical Center). The feasibility of collecting our predefined standard set was tested using data from these six centres.

For practical reasons, data were collected for a maximum of 100 consecutive patients per centre. In order to compare outcome trends between centres, centre-specific effects were examined for spirometry variables (forced vital capacity (FVC) % predicted, forced expiratory volume in 1 s (FEV₁) % pred), diffusing capacity of the lung for carbon monoxide (D_{LCO}) % pred and body mass index (BMI).

Some centres collected less data and were unable to include 100 patients. This was due to the fact the total number of patients that fit the criteria seen within the defined time frame was not always met.

In each centre, one or two pulmonologists were involved in the study. They participated in webinars and joined meetings during international conferences to discuss data collection, resulting data and data analyses. These multiple meetings guided the discussion around the implications of the data.

Patient population

The standard set was designed for patients with pulmonary sarcoidosis, with or without extrapulmonary disease [9, 10]. Patients with extrapulmonary involvement only were excluded.

Affiliations: ¹Dept of Value-Based Healthcare, St Antonius Hospital, Nieuwegein, The Netherlands. ²Radboud University Medical Center, Radboud Institute for Health Sciences, Scientific Center for Quality of Healthcare (IQ Healthcare), Nijmegen, The Netherlands. ³Interstitial Lung Diseases Center of Excellence, Dept of Pulmonology, St Antonius Hospital, Nieuwegein, The Netherlands. ⁴Dept of Pulmonary Medicine, Cleveland Clinic, Cleveland, OH, USA. ⁵Dept of Medicine, University of Cincinnati Medical Center, Cincinnati, OH, USA. ⁶Interstitial Lung Disease Unit, Royal Brompton Hospital, Imperial College, London, UK. ⁷Dept of Respiratory Medicine, Unit for Interstitial Lung Diseases, University Hospitals Leuven, Leuven, Belgium. ⁸Division of Heart and Lungs, University Medical Center Utrecht, Utrecht, The Netherlands. ⁹Dept of Pulmonary Medicine, Erasmus MC, University Medical Center, Rotterdam, The Netherlands. ¹⁰Dept of Biostatistics and Research Support, Julius Center for Health Sciences and Primary Care, Utrecht, The Netherlands. ¹¹Dept of Internal Medicine, University Medical Centre Utrecht, Utrecht, The Netherlands.

Correspondence: Nynke A. Kampstra, St Antonius Hospital, Koekoekslaan 1, 3435 CM, Nieuwegein, The Netherlands. E-mail: nynkekampstra@gmail.com

Study population: inclusion and exclusion criteria

We selected patients in line with the following criteria. Patients diagnosed between January 2010 and August 2014; follow-up time after diagnosis was ≥ 2 years; when clinical, radiological and/or bronchoalveolar lavage findings were compatible with the diagnosis, a tissue biopsy was optional; and the definition of pulmonary sarcoidosis patients was in line with the internationally accepted statement on sarcoidosis [11]: 1) the patient has to be diagnosed with pulmonary sarcoidosis, 2) the diagnosis is performed by a pulmonologist.

Patients were excluded when >2 years elapsed between initial diagnosis and first clinic visit and when <2 years of pulmonary function follow-up and treatment had been provided outside of the participating centre(s).

Outcome measures

Outcome measures were organised per domain, also referred to as a tier by PORTER and co-workers [2, 3]. The three-tiered hierarchy covers different domains of care received by patients during treatment. Tier 1 indicates the health status (survival and degree of health); tier 2 represents the recovery process (time to recovery, disutility of care); and tier 3 covers the sustainability of health (*i.e.* sustainability of health or recovery and long-term consequences of therapy). Our predefined standard set for pulmonary sarcoidosis includes seven outcome measures. Under tier 1, 1) mortality, 2) changes in pulmonary function (FVC, FEV₁, D_{LCO}) and 3) soluble interleukin-2 receptor (sIL-2R) change as an activity biomarker. Under tier 2 we defined one outcome measure: 4) weight changes. Finally, under tier 3, we defined three outcome measures: 5) quality of life (QoL), 6) osteoporosis and 7) clinical outcome status (COS). For osteoporosis, different protocols exist per centre. We were unable to evaluate the total number of patients assessed for the risk of developing osteoporosis. We were only able to identify the number of patients who had a dual energy X-ray absorptiometry (DEXA) scan during the respective follow-up period. The COS summarises different clinical phenotypes of the disease into nine categories [8]. The full background on how this set was developed is described elsewhere [8].

Pulmonary function tests were performed using Master Screen Body (Jaeger MS-PFT Analyser Unit, Würzburg, Germany), except for Cincinnati Medical Center, where the nSpire machine (nSpire KoKo Legend, model 315002, Longmont, CO, USA) was used. Changes in percentage predicted FVC, FEV₁ and D_{LCO} were measured. European Respiratory Society prediction equations for pulmonary function were used by all centres [12]. Changes in pulmonary function over time were then compared according to ethnicity, sex, treatment and centre effect.

Feasibility

Feasibility is defined as the “extent to which the required data are readily available or could be captured without undue burden and can be implemented for performance measurement” [13]. More specifically, the definition of feasibility in this study was evaluated by estimating, for each outcome measure, the percentage of patients for whom the outcome measure was assessed, out of the total number of patients. An outcome measure was considered feasible if collected on average in $\geq 80\%$ of the patients across centres.

Data collection

All consecutive patients diagnosed between January 2010 and August 2014 were eligible for this study. Data collection took place between May 2015 and May 2017, resulting in a maximum follow-up time of 96 months. By definition, patients who died within the 2-year follow-up after the diagnosis period were excluded. Data for a maximum of 100 patients were collected per centre. A REDCap database (version 8.1.11; www.project-redcap.org) was constructed by the study coordinator. Every centre assigned a data manager to enter outcome measures into REDCap. All required data were extracted manually from the electronic patient records.

Statistical analysis

All categorical data are presented in n (%) of the total patient group from the respective centre. Data on continuous variables are presented as mean \pm SD.

A linear mixed model with a random intercept, random slope was used to estimate trajectories over time for the BMI, and for percentage predicted FEV₁, FVC and D_{LCO}. The linear mixed model comprised centre, time and the interaction between centre and time. The likelihood ratio test was used to evaluate whether insertion of a variable improved the fit of the model. All analyses were performed in SPSS (version 24; IBM, Armonk, NY, USA).

Results

Data from 509 patients were analysed in this multicentre cohort, including patients from the St Antonius Hospital (n=100), Cincinnati Medical Center (n=55), Cleveland Clinic (n=97), Erasmus Medical Center (n=70), Leuven University Hospital (n=90) and Royal Brompton Hospital (n=97). Feasibility of data collection is shown in table 1. For each centre, baseline characteristics are presented in table 2. Mean age at diagnosis varied between 44.3±11.1 years (St Antonius hospital) and 52.1±12.3 years (Cincinnati Medical Center). At the Royal Brompton Hospital, a significantly higher proportion of patients had a Scadding stage IV chest radiograph (n=30, 30.9%), as reflected in low start of the intercept line in the linear mixed model, reflecting worse pulmonary function (figure 1a–c). The outcomes of the standard set are presented in table 3, except for lung function outcomes (figure 1a–c) and BMI trends (figure 2).

Tier 1: outcome measures 1, 2 and 3 (degree of health)

Mortality

As defined in the criteria, patients who died within the 2-year follow-up after diagnosis were excluded, so no 2-year mortality can be indicated. In our overall follow-up, six patients died from three different centres; in table 3 the mean number of months between the diagnosis and when the patient died are reported. Feasibility analysis showed that this variable was collected for 100% of the patients (table 1).

Pulmonary function

The results of the linear mixed model analyses for pulmonary function are presented in figure 1a–c. There is a yearly average increase of 0.8% for FVC (95% CI 0.4–1.1%), 0.7% for FEV₁ (95% CI 0.4–1.1%) and 0.3% for D_{LCO} (95% CI –0.1–0.7%). The patients from the centre with the highest FVC started at 111 (95% CI 97–125)% pred and the patients from the centre with the lowest started at 99 (95% CI 92–108)% pred. Baseline FEV₁ ranged from 98% (95% CI 84–112%) to 86% (95% CI 78–96%). No interaction effect between centre and time was observed for average change in FVC, FEV₁ and D_{LCO} over a period of 2 years. Although initial pulmonary function differed between centres, the changes in pulmonary function over time did not. Thus, none of the centres differed in pulmonary function test (PFT) trends over time compared to the other centres.

Table 4 indicates that for 69.2–70.1% of the patients, the change from baseline was within –10% and +10% change. There was no difference between centres in the proportion of patients with >10% change in PFT parameters. In addition, the older the patients, the more likely a deterioration of D_{LCO} would be observed (p=0.004). Feasibility analysis showed that this variable was collected for 100% of the patients (table 1).

sIL-2R

sIL-2R is only routinely measured during follow-up in the St Antonius Hospital. The sIL-2R levels (in pg·mL⁻¹) were measured in serum with enzyme immunoassays. Table 3 shows mean baseline sIL-2R levels of 87 patients at the St Antonius Hospital (5619±4099 pg·mL⁻¹). Baseline sIL-2R levels were significantly higher (8241±6344 pg·mL⁻¹, p<0.05) in patients ending up in COS 7–9 (n=53) 2 years after diagnosis

TABLE 1 Summary of the standard set for patients with pulmonary sarcoidosis

	Feasible # yes/no	Patients	Hospitals
Tier 1			
Mortality	Yes (100)	509/509	6/6
Pulmonary function	Yes (100)	509/509	6/6
sIL-2R change as an activity biomarker	No (16.5)	84/509	1/6 [¶]
Tier 2			
Weight gain	Yes (81)	412/509	5/6
Tier 3			
Quality of life; physical functioning	No (0)	0/0	0/6
Osteoporosis	NM	NM	6/6 ⁺
COS	Yes (81)	412/509	5/6

Data are presented as n or n (%), unless otherwise stated. sIL-2R: soluble interleukin-2 receptor; COS: clinical outcome status; NM: not measurable. #: the outcome measure was considered feasible when this was measured in >80% of all patients; [¶]: sIL-2R was feasible in St Antonius Hospital (it was measured in 84% of the patients), but in the other centres it was not routinely measured; ⁺: all centres make use of a local protocol to prevent osteopenia/osteoporosis in high-risk patients (in patients with long-term and/or high-dose prednisone use).

TABLE 2 Baseline characteristics of the multicentre study cohort

	St Antonius Hospital	Cincinnati Medical Center	Cleveland Clinic	Erasmus Medical Center	UZ Leuven	Royal Brompton Hospital
Subjects	100	55	97	70	90	97
Age at diagnosis years	44.3±11.1	52.1±12.3	47.6±10.5	44.7±11.4	45.4±11.4	48.3±12.5
Male	56 (56.0)	16 (29.1)	41 (42.3)	43 (61.4)	52 (57.8)	50 (51.5) [¶]
Ethnicity[#]						
White	79 (79.0)	44 (80.0)	70 (72.1)	54 (77.1)	82 (91.1)	71 (73.2)
Black	7 (7.0)	11 (20.0)	25 (25.8)	11 (15.7)	1 (1.1)	0
Other	14 (14.0)	0	0	4 (5.7)	5 (5.6)	26 (26.8)
Unknown	0	0	2 (2.1)	1 (1.5)	2 (2.2)	0
Weight at diagnosis kg	83.0±17.0	88.5±24.0	86.9±20.7	83.4±18.4	86.0±22.4	82±21.9
BMI at diagnosis kg·m⁻²	26.8±4.8	31.8±8.1	31.2±8.9	27.6±5.6	28.0±5.3	29.0±7.7
Smoking						
Never	51 (51.0)	32 (58.2)	57 (58.8)	35 (50.0)	60 (66.7)	61 (62.9)
Ever	33 (33.0)	18 (32.7)	36 (37.1)	24 (34.3)	22 (24.4)	30 (30.9)
Active	11 (11.0)	3 (5.5)	4 (4.1)	11 (15.7)	8 (8.9)	6 (6.2)
Unknown	5 (5.0)	2 (3.6)	0	0	0	0
Scadding stage						
Stage 0	4 (4.0)	1 (1.8)	6 (6.2)	9 (12.9)	18 (20)	6 (6.2)
Stage I	36 (36.0)	24 (43.6)	23 (23.7)	20 (28.6)	37 (41.1)	20 (20.6)
Stage II or III	48 (48.0)	16 (29.1)	41 (42.3)	33 (47.1)	27 (30.0)	39 (40.2)
Stage IV	11 (11.0)	5 (9.1)	8 (8.2)	0	1 (1.1)	30 (30.9)
Unknown	1 (1.0)	9 (16.4)	19 (19.6)	8 (11.4)	7 (7.8)	2 (2.1)
Mean 1st PFT						
FVC % pred	97.2±17.6	89.8±26.4	94.8±18.4	96.1±19.0	99.4±17.7	87.3±23.4
FEV ₁ % pred	89.9±18.2	82.2±25.3	86.0±19.4	87.0±23.4	91.7±19.6	77.8±22.4
D _{LCO} % pred	77.3±16.4	77.8±22.0	77.7±16.3	75.1±14.8	74.8±16.3	64.8±20.0

Data are presented as n, mean±SD or n (%) of the total patient group from each clinic. n=509. BMI: body mass index; PFT: pulmonary function test; FVC: forced vital capacity; FEV₁: forced expiratory volume in 1 s; D_{LCO}: diffusing capacity of the lung for carbon monoxide. [#]: ethnicity was collected in the categories black, white or Asian; however, in the Royal Brompton, this was collected as Caucasian or non-Caucasian; [¶]: one subject's sex was unknown.

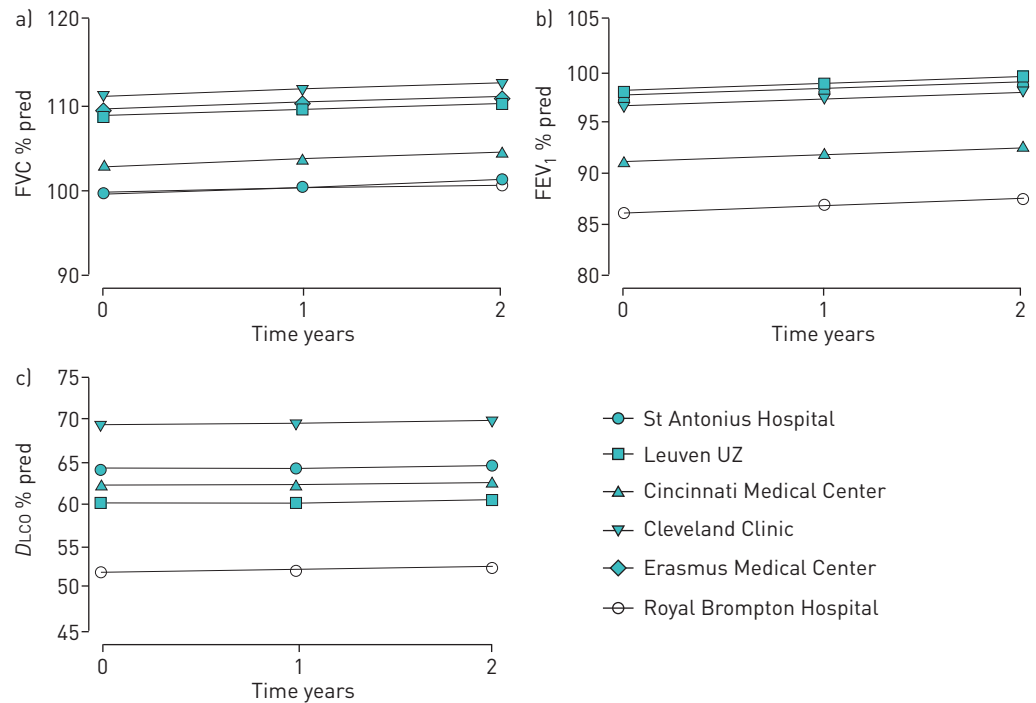


FIGURE 1 Linear mixed model pulmonary function parameters over time. All pulmonary function results are based on the European Community for Steel and Coal reference equations. In a), the St Antonius Hospital and Leuven UZ and the Erasmus Medical Center (MC) and Royal Brompton Hospital follow a similar trend. In b), the St Antonius Hospital, Erasmus MC and Leuven UZ follow a similar trend. In c), the St Antonius hospital and the Erasmus MC follow a similar trend. Predictions are plotted for white patients. For non-white patients, the intercept (baseline) starts lower: a) -16.8 , b) -12.6 and c) -11.4 for each centre. The trend follows the same line as presented above. Linear-mixed model with fixed effects: time in study and ethnicity.

compared to patients ending up in COS 1–6 ($n=33$, 6232 ± 2438 $\text{pg}\cdot\text{mL}^{-1}$). For the St Antonius Hospital alone, it was a feasible outcome measure as it was collected for 84% of patients; however, data collection for this variable overall was not feasible (measured in 16.5% of the patients) (table 1).

Tier 2: outcome measure 4 (disutility of care)

Weight changes

Results of the linear mixed model analyses for weight changes are presented in figure 2. There is an overall increase of 0.1 (95% CI 0.03–0.21) $\text{kg}\cdot\text{m}^{-2}$ in BMI per year for this specific outcome, which includes data from five out of the six centres. There was no centre–time interaction, *i.e.* no significant differences were seen between centres in BMI changes over time. However, baseline BMI values differed between centres: baseline BMI from Cincinnati Medical Center was 31.8 (95% CI 28.1–35.6) $\text{kg}\cdot\text{m}^{-2}$, while it was 26.8 (95% CI 23.6–29.4) $\text{kg}\cdot\text{m}^{-2}$ in the St Antonius Hospital cohort. It was feasible to collect data on weight changes, as it was collected for 81% of the patients (table 1).

Tier 3: outcome measure 5, 6 and 7 (long-term consequences of therapy)

Health-related QoL

For this outcome, no structured multicentre data were collected. Health-related QoL was measured regularly in a minority of patients if they were participating in a clinical trial. In addition, some centres only measured QoL once at baseline, when the patient visited the clinic for the first time. In this retrospective study, it was not feasible to collect this outcome measure routinely (table 1).

Osteoporosis

Few patients were diagnosed with osteoporosis or osteopenia (table 3). 11 (0.6%) out of 509 patients were diagnosed with osteoporosis and six (1.6%) out of 509 patients were diagnosed with osteopenia. Another six patients in whom a DEXA scan was performed had a normal T-score. It was not possible to provide an estimation on the feasibility for this outcome measure, for a number of reasons. First, it was not possible to identify those patients at risk of developing osteoporosis. Second, it was unclear which patients

TABLE 3 Outcome measures of the multicentre study cohort

	St Antonius Hospital	Cincinnati Medical Center	Cleveland Clinic	Erasmus Medical Center	Leuven UZ	Royal Brompton Hospital
Subjects n	100	55	97	83	90	97
Tier 1						
Mortality	0	1	0	0	1	4
Months between diagnosis and date when patient died	0	56	0	0	42	32.3±27.6
Mean number of PFTs sIL-2R	5.0±3.0	1.7±0.9	3.6±2.5	4.6±3.3	5.4±4.2	3.0±1.6
Baseline sIL-2R pg·mL ⁻¹ (n=87)	5619±4099	ND	ND	ND	ND	ND
Tier 2						
Weight gain BMI changes	0.7	0.4	0.03	0.2	-0.1	0.2
Number of months between first/last BMI	36.8±10.8	15.6±16.9	36.7±16.8	41.4±15.5	34.4±23.7	24.4±8.0
Tier 3						
Quality of life	ND	ND	ND	ND	ND	ND
Osteoporosis						
Osteopenia [#]	3 (3.0)	1 (1.8)	1 (1.0)	0	1 (1.1)	0
Osteoporosis [†]	2 (2.0)	3 (5.5)	5 (5.2)	1 (1.2)	0	0
Normal T-score	4 (4.0)	0	0	2 (2.4)	0	0
COS ⁺						
COS 1-3	18 (18.0)	2 (3.6)	16 (16.5)	7 (8.4)	14 (15.5)	3 (3.1)
COS 4-6	26 (26.0)	9 (16.4)	10 (10.3)	25 (30.1)	28 (31.0)	36 (37.1)
COS 7-9	53 (53.0)	44 (80.0)	71 (73.2)	51 (61.5)	37 (41.2)	54 (55.7)
Unknown	3 (3.0)	0	0	0	11 (12.3)	4 (4.1)

Data are presented as n, mean±SD or n (%) of the total patient group from each clinic. n=509. Pulmonary function test (PFT) (outcome measure 2) changes are presented in figure 1a-c, as a linear mixed model was used to assess changes. sIL-2R: soluble interleukin-2 receptor; BMI: body mass index; COS: clinical outcome status; ND: no data. #: osteopenia T-score: between -1.0 and -2.5; †: osteoporosis T-score ≤-2.5; +: COS: 1=resolved never treated, 2=resolved no therapy >1 year, 3=minimal disease never treated, 4=minimal disease no therapy >1 year, 5=persistent no current therapy never treated, 6=persistent no current therapy no therapy >1 year, 7=persistent current therapy asymptomatic, 8=persistent current therapy symptomatic, 9=persistent current therapy worsening prior year.

were monitored according to a locally available protocol on their bone health status. Third, we were unable to identify whether there was an active decision by the pulmonologist to not perform a DEXA scan.

COS

COS was evaluated in all centres (table 3). Overall, out of the 509 patients where COS was evaluated, 310 (60.9%) patients were still requiring systemic therapy 2 years after the diagnosis. In this retrospective cohort, it was feasible to collect data internationally for this outcome measure (collected in 100% of patients).

Discussion

In this study we present the results of the first evaluation of the standard set, previously developed for pulmonary sarcoidosis patients. We show that collecting data in a retrospective multicentre international cohort is feasible for the outcomes mortality, pulmonary function, weight changes and COS. Conversely, it is not (yet) feasible to collect data for international comparisons on QoL, sIL-2R and osteoporosis.

We developed a standard set for pulmonary sarcoidosis patients in 2015 [8]. Until now, to our knowledge, there have been no scientific papers implementing and reflecting on a standard set for pulmonary sarcoidosis patients in a clinical setting. Although many standard sets for other conditions have been developed recently, real-world implementation of outcome measures and how they function in practice are lacking [1]. Our study provides information on the feasibility of implementing the pulmonary sarcoidosis standard set. We present the first results of a full standard set developed for pulmonary sarcoidosis patients in collaboration with six ILD centres from three different countries. Recruitment of the centres focused on specialised sarcoidosis centres, which potentially have differences in baseline clinical features and ethnicity. In addition, treatment was different between centres. Although several ILD clinics report on similar clinical outcomes such as spirometry and weight changes, globally there is little consensus which clinical outcomes should be reported on [14]. In order to realise the implementation of VBHC, it is

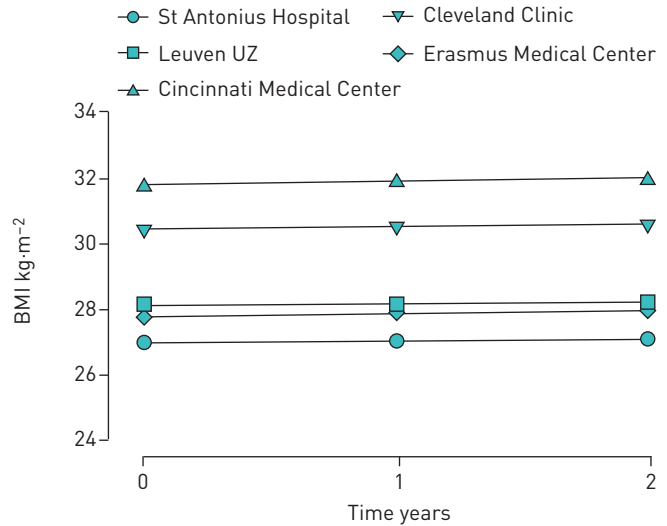


FIGURE 2 Linear mixed model body mass index (BMI) over time. Linear mixed model with fixed effects: time in study. The Royal Brompton Hospital was not able to collect weight changes over time.

important to know what the most important outcomes and their feasibility are to report on for RCTs as well as for quality of care oriented studies.

Based on our experience with the standard set for pulmonary sarcoidosis, it is feasible to implement some outcomes of the standard set in clinical practice. We do wish to critically reflect on the outcome measures and the potential to use the results as a basis for improving the quality of care.

Outcomes measures for pulmonary sarcoidosis

Tier 1: degree of health

Mortality

Mortality for pulmonary sarcoidosis is generally considered to be low, although there are subsets of patients, mainly those with extensive fibrotic lung disease and those with significant cardiac sarcoidosis, who experience markedly higher mortality rates than the general population [15]. Compared to Scadding stages I–III, patients with stage IV patients have a worse survival [16], and evidence of fibrosis on chest computed tomography is an independent risk factor for higher mortality [15, 17].

It may be more meaningful and clinically relevant to use mortality as an outcome measure in the subgroup with pulmonary fibrosis on imaging [16, 17]. Indeed, the highest number of deaths were observed at the Royal Brompton Hospital, where a significantly higher proportion of patients with stage IV disease were observed compared to the other centres. Especially for clinical trials, measuring mortality is feasible although not always a clinically meaningful and relevant outcome in pulmonary sarcoidosis patients. In a subset of patients this might be more meaningful, such as in pulmonary fibrosis.

Lung function

As expected, due to variation in the patient composition of the cohorts, the six participating centres varied in baseline FVC. However, we did not find variation in the lung function trends over a period of 2 years between the six centres. Pulmonary sarcoidosis may cause significant airflow obstruction. However, we did not find significant differences between centres in FEV₁ trend over time [18]. Other ILDs, such as idiopathic pulmonary fibrosis show a more progressive disease pattern with FVC worsening and high mortality rates over time. Nevertheless, pulmonary sarcoidosis is progressive in a subgroup of patients. We observed a tendency for an improvement over time for the three spirometry parameters. In a large (n=1774) cohort consisting of sarcoidosis patients, pulmonary function improved over time [19]. FVC has been recommended as the first-choice clinical end-point for pulmonary sarcoidosis trials [20], but FVC does not always accurately reflect the varying disease-related manifestations [14]. Looking at our own data, this outcome measure was not able to identify potential practice variation. In addition, over time FVC showed minor improvements. Therefore, it could be reconsidered whether this should be the outcome measure of first choice, also from the patient's perspective. A responder analysis could be useful in identifying the proportion of patients who benefit and in defining a responder cut-off.

TABLE 4 Pulmonary function changes between first and last measurement

	≥−10% worsening	−10–10% change	≥10% improvement	p-value
FVC % pred				
Subjects (n=509)	50 (9.8)	352 (69.2)	107 (21.0)	
Mean FVC change	−16.6±6.9	0.7±4.8	18.4±8.5	
Time between first and last PFT months	37.0±16.8	31.1±17.7	34.4±17.4	0.035
Age at diagnosis years	48.9±10.6	46.9±11.7	45.5±12.2	0.231
Weight kg	88.4±20.7	85.4±19.6	83.6±22.7	0.399
FEV₁ % pred				
Subjects (n=509)	44 (8.7)	357 (70.1)	108 (21.2)	
Mean FEV ₁ change	−18.5±7.8	0.4±4.9	18.9±12.7	
Time between first and last PFT months	40.5±16.8	31.0±17.6	33.7±17.3	0.002
Age at diagnosis years	48.3±11.6	47.0±11.3	45.6±12.7	0.383
Weight kg	91.0±19.3	84.2±20.1	86.5±21.5	0.100
D_{LCO} % pred				
Subjects (n=492)	67 (13.6)	341 (69.3)	84 (17.1)	
Mean D _{LCO} change	−15.6±5.4	0.4±5.0	17.0±5.4	
Time between first and last PFT months	35.2±16.7	31.6±16.9	36.6±18.4	0.029
Age at diagnosis years	49.5±11.8	47.1±11.4	43.4±12.0	0.004
Weight kg	86.3±20.1	84.8±19.8	86.2±23.7	0.769

Data are presented as n (%) or mean±SD, unless otherwise stated. p-values are based on ANOVA test. FVC: forced vital capacity; PFT: pulmonary function test; FEV₁: forced expiratory volume in 1 s; D_{LCO}: diffusing capacity of the lung for carbon monoxide.

sIL-2R

Based on analysis of the retrospective cohort consisting of 87 patients treated at the St Antonius Hospital, this biomarker was considered a possible feasible outcome measure in specific settings. In addition, it could potentially be used as a predictive biomarker, as we have shown that patients with elevated baseline sIL-2R levels are at risk of requiring therapy 2 years after the diagnosis. This may be confounded, as the treating physician may be less inclined to discontinue therapy in someone with higher baseline sIL-2R. Based on a review it was concluded that there are conflicting data concerning the correlation between changes in sIL-2R level and treatment effects such as PFT [21]. Therefore, the role of determining sIL-2R in clinical practice needs further research.

Looking at the total multicentre retrospective cohort, it is not feasible to collect data of this outcome measure. In the Dutch setting this might be feasible, as sIL-2R data was collected in 84% of the patients. This is also related to higher costs of the sIL-2R measurements in some of the centres.

Tier 2: disutility of care

BMI was a feasible parameter to collect (81%). No differences between centres were observed in BMI trends over time. It may be that with more prolonged periods of observation differences in trend over time would become apparent. BMI trends over time did not appear to differ between centres. However, for baseline BMI, we did see substantial differences. As weight gain is an important side-effect, often due to high-dose and long-term glucocorticoid use, this can be an important outcome measure to achieve improved quality of care in the future [22]. Additionally, prolonged high-dose glucocorticoid therapy is associated with a large number of negative side-effects, including weight gain, the development of osteoporosis and diabetes [22].

Tier 3

COS

Out of 509 patients, 310 (60.9%) patients were still requiring systemic therapy 2 years after the diagnosis. However, it was challenging to collect the COS data, as this is not a routinely reported outcome measure in electronic patient records. Pulmonologist and data managers had to read the patients' electronic record and make a judgement on the COS. Medical history data used to judge the patients' COS could create some individual variability in interpretation. When simplifying COS, there may be less interobserver

variability in scoring the patients' COS. We therefore advise to simplify COS into three categories: 1) never treated (COS 1–3), 2) previously treated (COS 4–6) and 3) current therapy (COS 7–9). These main categories could also be used when performing a subgroup analysis.

QoL

Unfortunately, QoL was not yet routinely measured in any of the centres. Therefore, we are not able to evaluate the added value of collecting and comparing this outcome measure. Should centres wish to prioritise the implementation of patient-reported outcome measures, additional staffing is required, facilitating data collection.

Osteoporosis and osteopenia

Glucocorticoids are considered first-line therapy for treating pulmonary sarcoidosis. Sarcoidosis patients may be treated with glucocorticoids for a long time, sometimes even for years. Long-term use of glucocorticoids can eventually increase the risk of developing osteoporosis [23, 24]. Untreated sarcoidosis has also been associated with low bone mineral density [25]. Moreover, corticosteroid use can cause increased bone resorption, decreased bone formation and lead to a net bone loss. In our retrospective cohort few patients were diagnosed with osteopenia or osteoporosis. We believe this percentage may be too low and it is likely that this outcome was under-reported.

A significant dose–response relationship was observed between the cumulative dose of glucocorticoids and risk of any fracture [26]. In a retrospective cohort study of newly diagnosed sarcoidosis patients consisting of 105 patients ever treated with glucocorticoids (GC group), and 49 patients (no-GC group) not treated with glucocorticoids during a follow-up of 101 months, adverse consequences due to glucocorticoids therapy were evaluated [22]. In the GC group, 16 (15.2%) patients developed new osteoporosis or osteopenia. In the no-GC group, five (10.2%) patients developed osteoporosis or osteopenia. Therefore, this outcome measure is mainly of importance for long-term and high-dose glucocorticoids users.

Another reason for low numbers of osteoporosis and osteopenia can be due to the fact that referral centres performed a DEXA scan, but did not share the results. We were unable to evaluate the feasibility of this outcome measure. As earlier stressed, we were unable to identify those patients being at risk of developing osteopenia/osteoporosis in each centre. It would be useful to develop an international standard for when to perform a DEXA scan.

Organisational level

Healthcare systems are complex and there are large differences between healthcare systems globally, which can be a barrier to having valid international comparison of outcomes [1]. In order to have solid ongoing data collection, it is crucial to have a dedicated person in each centre to manage the data collection process. Secondly, it is important that the team members have good understanding of how the outcome measures can be used. When there is good understanding of the potential of outcome measures, this can be used more effectively for quality improvement efforts or for the planning of future care [27].

The objectives of clinical reporting include increasing (shared) decision making for patients and physicians, encourage quality improvement, better performance management and contributing to research [28]. When data and outcome measures are made public, this may promote improvement in care, as centres are more likely to discuss results and change practices. Thus, transparency can make physicians aware of practice variations and may help to improve care.

Limitations

First, although we were able to present data in an international multicentre initiative, our initiative represents only centres from the United States and three European countries. There was no representation from Asia, Australasia, Africa or South America. International involvement is key in order to have good implementation of the standard outcome set. Hopefully in the future the standard set can be evaluated and validated as part of a larger international network of ILD clinics, for example, in a yearly evaluation cycle. However, for the Netherlands and the USA, we have reached a reasonable representation of ILD expert clinics. Two out of the three ILD expert centres in the Netherlands (officially declared by the Ministry of Health, Welfare and Sport) were part of the expert group.

Second, data were collected retrospectively. Patients visited the clinic at different time points, which makes it more difficult to evaluate and compare data. Furthermore, the centres measure and report some outcomes differently, such as ethnicity. In some centres ethnicity can only be evaluated based on photographs in the electronic patient record. Other centres report ethnicity only in two categories: Caucasian or non-Caucasian or Hispanic or non-Hispanic. Third, we were required to extract most of the data from the hospital databases manually, as an automated system for extracting these data in the

respective centres does not currently exist. It was challenging to collect some of the required variables of the standard set, for example the COS data, as this information is not routinely reported. Fourth, underlying variation in routine daily clinical practice may exist due to differences in care pathways, but also due to costs between countries when measuring certain clinical outcome measures. Fifth, the definition of feasibility is strongly influenced by the number of patients being part of the cohort from each centre. Sixth, we were unable to integrate QoL outcomes. The potential of routinely measuring QoL in daily clinical practice can be an outcome measure of great importance. It would be of great value if future trials more closely reflected on how systematically measuring QoL can enable improvements in the quality of care. Although some centres measure this, QoL is rarely being followed-up on a systematic basis.

Finally, we do not know whether this standard set is the most relevant from a patient's perspective. As stressed earlier, FVC is considered important for the follow-up of sarcoidosis. These are chosen for convenience and feasibility rather than this being the optimal standard set from a patient's perspective. Therefore, evaluating what patients consider to be the most important and relevant outcomes is key for future studies.

Conclusions

We have developed a standard set that can be used both retrospectively and prospectively to compare clinical outcomes of different sarcoidosis centres. The dataset can be improved by collecting this information prospectively over time. This set can support further implementation of VBHC and the standardisation of collecting outcomes for pulmonary sarcoidosis patients. Efforts on systematically measuring outcomes have the potential to create opportunities for improving treatment decision making and realise benchmarking between different healthcare systems, both on a regional and international level.

In conclusion, for international collection of a standard outcome set in pulmonary sarcoidosis, based on our definition that it was measured in >80% of the patients, this study showed that it is feasible to collect four out of seven outcome measures; data on mortality, changes in pulmonary function, changes in weight/BMI and the COS. It is not (yet) feasible to retrospectively collect data on sIL-2R, osteoporosis and QoL internationally.

Acknowledgements: We thank the sarcoidosis patient association (SBN) for the research grant awarded in 2016 for this project. We would like to thank Bernt van den Blink for his helpful input during the start of this study, during which he worked as a pulmonologist at the Erasmus Medical Center (Rotterdam, the Netherlands). We would like to thank the students and data managers from the respective six centres who entered data into REDCap.

Support statement: This work was supported by The Netherlands Organisation for Health Research and Development (ZonMw) under project number 842001005. The funder had no role in the study design, data collection, analysis or decision of where to publish the manuscript. Funding information for this article has been deposited with the Crossref Funder Registry.

Author contributions: All authors contributed to the conception and design of the study. N.A. Kampstra wrote the first draft of the manuscript. N.A. Kampstra, M.J.C. Eijkemans, L.M. Dijkman and P. Zanen analysed and interpreted the data. N.A. Kampstra and P. Zanen had full access to all data in the study, and take responsibility for the integrity of the data and accuracy of the data analysis. All authors contributed to the analysis and interpretation, and contributed to the drafting of the article. All authors contributed to the revision of the article for important intellectual content. All authors read and approved the final manuscript.

Conflict of interest: N.A. Kampstra has nothing to disclose. P.B. Van der Nat has nothing to disclose. L.M. Dijkman has nothing to disclose. F.T. van Beek has nothing to disclose. D.A. Culver reports nonfinancial support from Gilead, grants and other support from Mallinkrodt, and nonfinancial support from Araim, outside the submitted work. R.P. Baughman has nothing to disclose. E.A. Renzoni reports personal fees from Roche and Boehringer outside the submitted work. W. Wuyts reports grants from Roche and Boehringer Ingelheim paid to his institution outside the submitted work. V. Kouranos has nothing to disclose. P. Zanen has nothing to disclose. M.S. Wijsenbeek has nothing to disclose. M.J.C. Eijkemans has nothing to disclose. D.H. Biesma has nothing to disclose. P.J. van der Wees has nothing to disclose. J.C. Grutters reports grants from ZonMw during the conduct of the study.

References

- 1 Michelotti M, de Korne DF, Weizer JS, *et al.* Mapping standard ophthalmic outcome sets to metrics currently reported in eight eye hospitals. *BMC Ophthalmol* 2017; 17: 269.
- 2 Porter ME. What is value in health care? *N Engl J Med* 2010; 363: 2477–2481.
- 3 Porter ME, Larsson S, Lee TH. Standardizing patient outcomes measurement. *N Engl J Med* 2016; 374: 504–506.
- 4 Forde I, Morgan D, Klazinga NS. Resolving the challenges in the international comparison of health systems: the must do's and the trade-offs. *Health Policy* 2013; 112: 4–8.
- 5 Kim AH, Roberts C, Feagan BG, *et al.* Developing a standard set of patient-centred outcomes for inflammatory bowel disease – an international, cross-disciplinary consensus. *J Crohns Colitis* 2018; 12: 408–418.
- 6 Daeter EJ, Timmermans MJ, Hirsch A, *et al.* Defining and measuring a standard set of patient-relevant outcomes in coronary artery disease. *Am J Cardiol* 2018; 121: 1477–1488.
- 7 Baughman RP, Nagai S, Balter M, *et al.* Defining the clinical outcome status (COS) in sarcoidosis: results of WASOG Task Force. *Sarcoidosis Vasc Diffuse Lung Dis* 2011; 28: 56–64.

- 8 Kampstra NA, Grutters JC, van Beek FT, *et al.* A first patient-centred set of outcomes for pulmonary sarcoidosis: a multicentre initiative. *BMJ Open Respir Res* 2019; 6: e000394.
- 9 Iannuzzi MC, Fontana JR. Sarcoidosis: clinical presentation, immunopathogenesis, and therapeutics. *JAMA* 2011; 305: 391–399.
- 10 Baughman RP, Lower EE, Gibson K. Pulmonary manifestations of sarcoidosis. *Presse Med* 2012; 41: e289–e302.
- 11 Costabel U, Hunninghake G. ATS/ERS/WASOG statement on sarcoidosis. *Eur Respir J* 1999; 14: 735–737.
- 12 Qunjer PH. Standardized lung function testing. *B Eur Physiopath Res* 1983; 19: 39–44.
- 13 National Quality Forum. Report from the National Quality Forum: eMeasure Feasibility Assessment. April 2013. www.qualityforum.org/Publications/2013/04/eMeasure_Feasibility_Assessment.aspx
- 14 Judson MA. Endpoints in sarcoidosis: more like IPF or asthma? *Respir Med* 2018; 138S: S3–S4.
- 15 Kouranos V, Wells A, Walsh S. Why do people die from pulmonary sarcoidosis? *Curr Opin Pulm Med* 2018; 24: 527–535.
- 16 Nardi A, Brillet PY, Letoumelin P, *et al.* Stage IV sarcoidosis: comparison of survival with the general population and causes of death. *Eur Respir J* 2011; 38: 1368–1373.
- 17 Kirkil G, Lower EE, Baughman RP. Predictors of mortality in pulmonary sarcoidosis. *Chest* 2018; 153: 105–113.
- 18 Polychronopoulos VS, Prakash UB. Airway involvement in sarcoidosis. *Chest* 2009; 136: 1371–1380.
- 19 Judson M, Boan A, Lackland D. The clinical course of sarcoidosis: presentation, diagnosis, and treatment in a large white and black cohort in the United States. *Sarcoidosis Vasc Diffuse Lung Dis* 2012; 29: 119–127.
- 20 Baughman R, Drent M, Culver D. Endpoints for clinical trials of sarcoidosis. *Sarcoidosis Vasc Diffuse Lung Dis* 2012; 29: 90–98.
- 21 Chopra A, Kalkanis A, Judson MA. Biomarkers in sarcoidosis. *Expert Rev Clin Immunol* 2016; 12: 1191–1208.
- 22 Khan NA, Donatelli CV, Tonelli AR, *et al.* Toxicity risk from glucocorticoids in sarcoidosis patients. *Respir Med* 2017; 132: 9–14.
- 23 Adler RA, Funkhouser HL, Petkov VI, *et al.* Glucocorticoid-induced osteoporosis in patients with sarcoidosis. *Am J Med Sci* 2003; 325: 1–6.
- 24 Grossman JM, Gordon R, Ranganath VK, *et al.* American College of Rheumatology 2010 recommendations for the prevention and treatment of glucocorticoid-induced osteoporosis. *Arthritis Care Res* 2010; 62: 1515–1526.
- 25 Rizzato G. Clinical impact of bone and calcium metabolism changes in sarcoidosis. *Thorax* 1998; 53: 425–429.
- 26 Balasubramanian A, Wade S, Adler R, *et al.* Glucocorticoid exposure and fracture risk in a cohort of US patients with selected conditions. *J Bone Miner Res* 2018; 33: 1881–1888.
- 27 Kampstra NA, Zipfel N, van der Nat PB, *et al.* Health outcomes measurement and organizational readiness support quality improvement: a systematic review. *BMC Health Serv Res* 2018; 18: 1005.
- 28 Mason A, Street A. Publishing outcome data: is it an effective approach? *J Eval Clin Pract* 2006; 12: 37–48.