Arthritis Care & Research
Vol. 75, No. 2, February 2023, pp 307–316
DOI 10.1002/acr.24785

American College of Rheumatology
Empowering Rheumatology Professionals

Clinical and Molecular Findings After Autologous Stem Cell Transplantation or Cyclophosphamide for Scleroderma: Handling Missing Longitudinal Data

Lynette Keyes-Elstein,[1] Ashley Pinckney,[1] Ellen Goldmuntz,[2] Beverly Welch,[2] Jennifer M. Franks,[3] Viktor Martyanov,[3] Tammara A. Wood,[3] Leslie Crofford,[4] Maureen Mayes,[5] Peter McSweeney,[6] Richard Nash,[6] George Georges,[7] M. E. Csuka,[8] Robert Simms,[3] Daniel Furst,[9] Dinesh Khanna,[10] E. William St.Clair,[11] Michael L. Whitfield,[3] and Keith M. Sullivan[11]

Objective. Among individuals with systemic sclerosis (SSc) randomized to cyclophosphamide (CYC) (n = 34) or hematopoietic stem cell transplantation (HSCT) (n = 33), we examined longitudinal trends of clinical, pulmonary function, and quality of life measures while accounting for the influence of early failures on treatment comparisons.

Methods. Assuming that data were missing at random, mixed-effects regression models were used to estimate longitudinal trends for clinical measures when comparing treatment groups. Results were compared to observed means and to longitudinal trends estimated from shared parameter models, assuming that data were missing not at random. Longitudinal trends for SSc intrinsic molecular subsets defined by baseline gene expression signatures (normal-like, inflammatory, and fibroproliferative signatures) were also studied.

Results. Available observed means for pulmonary function tests appeared to improve over time in both arms. However, after accounting for participant loss, forced vital capacity in HSCT recipients increased by 0.77 percentage points/year but worsened by –3.70/year for CYC ($P = 0.004$). Similar results were found for diffusing capacity for carbon monoxide and quality of life indicators. Results for both analytic models were consistent. HSCT recipients in the inflammatory (n = 20) and fibroproliferative (n = 20) subsets had superior long-term trends compared to CYC for pulmonary and quality of life measures. HSCT was also superior for modified Rodnan skin thickness scores in the fibroproliferative subset. For the normal-like subset (n = 22), superiority of HSCT was less apparent.

Conclusion. Longitudinal trends estimated from 2 statistical models affirm the efficacy of HSCT over CYC in severe SSc. Failure to account for early loss of participants may distort estimated clinical trends over the long term.

INTRODUCTION

Systemic Sclerosis (SSc; scleroderma) remains a devastating autoimmune disorder with mortality unchanged over the past 40 years (1–3). The Scleroderma: Cyclophosphamide or Transplantation (SCOT) clinical trial compared myeloablation followed by CD34+ selected autologous hematopoietic stem cell transplant (HSCT) versus 12 monthly infusions of cyclophosphamide (CYC) (4). Participants were followed for up to 72 months. The primary end point was a global rank composite score

ClinicalTrials.gov identifier: NCT00114530.

Supported by the NIH (National Institute of Allergy and Infectious Diseases grants N01-AI-05419 and HHSN272201100025C to Duke University and grants N01-AI-25481, HHSN272200900057C, and 1UM2AI117870 to Rho [the statistical and clinical coordinating center]), the Scleroderma Research Foundation (grant to Dr. Whitfield), the Burroughs-Wellcome PUP Big Data in the Life Sciences Training Program, and the Dr. Ralph and Marian Falk Medical Research Trust (grant to Dr. Whitfield through Dartmouth College). Dr. Franks' work was supported by the NIH (grant BD2K-T32-5T32LM012204-03).

[1]Lynette Keyes-Elstein, DrPH, Ashley Pinckney, MS: Rho Federal Systems Division, Durham, North Carolina; [2]Ellen Goldmuntz, MD, PhD, Beverly Welch, RN, MSN: National Institute of Allergy and Infectious Diseases, Bethesda, Maryland; [3]Jennifer M. Franks, PhD, Viktor Martyanov, PhD, Tammara A. Wood, MS, Robert Simms, MD, Michael L. Whitfield, PhD: Geisel School of Medicine at Dartmouth, Hanover, New Hampshire; [4]Leslie Crofford, MD: Vanderbilt University School of Medicine, Nashville, Tennessee; [5]Maureen Mayes, MD, MPH: University of Texas McGovern Medical School, Houston; [6]Peter McSweeney, MD, Richard Nash, MD: Colorado Blood Cancer Institute, Denver; [7]George Georges, MD: Fred Hutchinson Cancer Research Center, Seattle, Washington; [8]M. E. Csuka, MD: Medical College of Wisconsin, Milwaukee; [9]Daniel Furst, MD: University of California, Los Angeles, University of Washington, Seattle, and University of Florence, Florence, Italy; [10]Dinesh Khanna, MD: University of Michigan, Ann Arbor; [11]E. William St.Clair, MD, Keith M. Sullivan, MD: Duke University Medical Center, Durham, North Carolina.

Author disclosures are available at https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Facr.24785&file=acr24785-sup-0001-Disclosureform.pdf.

SCOT trial data are accessible at www.immport.org for study SDY1039 (DOI: 10.21430/M3SM4LTLH).

Address correspondence via email to Keith M. Sullivan, MD, at keith.sullivan@duke.edu.

Submitted for publication April 9, 2021; accepted in revised form September 14, 2021.

SIGNIFICANCE & INNOVATIONS

- Missing data in rheumatology trials can distort results of randomized studies.
- In the Scleroderma: Cyclophosphamide or Transplantation (SCOT) trial, early failures in the cyclophosphamide arm appeared to bias longitudinal results of pulmonary and quality of life measures.
- Statistical corrections for longitudinal data missing at random (mixed-effects regression models) and missing not at random (shared parameter models) showed that over time, myeloablative autologous hematopoietic stem cell recipients improved clinical and molecular markers of disease.
- For severe scleroderma, stem cell transplant offers an emerging innovation.

(GRCS) at 54 months comparing each participant with every other on the basis of a hierarchy of disease features: death, survival without respiratory, renal, or cardiac failure, forced vital capacity (FVC), Health Assessment Questionnaire disability index (HAQ DI) score, and modified Rodnan skin thickness score (MRSS). For the primary end point in the intent-to-treat population, myeloablative HSCT led to superior outcome compared to CYC (4). In a companion mechanistic study, molecular signatures of SSc were evaluated at month 26 and found to return to normal after HSCT but not after CYC (5).

To evaluate the clinical relevance of the SCOT results and provide a deeper understanding of the patterns of change over time, we conducted longitudinal analyses for clinical, laboratory, and quality of life assessments including FVC, diffusing capacity for carbon monoxide (DLco), MRSS, HAQ DI, and Short Form 36 (SF-36) health survey physical and mental composite scores. In any trial, misleading longitudinal trends may result from missing participant data. If ≥1 study arms have excess early failures (as was true for CYC recipients in SCOT), subsequent serial observations may be biased because only surviving (healthier) participants are available for long-term comparison (6,7). To address this distortion, statistical approaches have been proposed for data considered missing at random or missing not at random (8–10). We report here the longitudinal trends for SCOT treatment groups with and without analytic corrections for missing data and show how missing data may bias longitudinal comparisons of measures of scleroderma over time.

We also explored longitudinal trends for subsets of SCOT participants defined by baseline gene expression signatures (11,12). Four molecular intrinsic subsets of SSc patients, called inflammatory, fibroproliferative, normal-like, and limited, have been identified and validated across multiple cohorts using skin samples (13–18) and have subsequently been found in other tissues, including in peripheral blood cells (19–21). SCOT participants fall into 3 of the 4 intrinsic subsets at baseline: inflammatory, fibroproliferative, and normal-like (12). Franks et al have recently reported that SCOT participants with a fibroproliferative molecular signature evidenced improved survival after HSCT compared to CYC (21). We hypothesized that other longitudinal responses might also correlate with baseline gene signatures.

PATIENTS AND METHODS

Study participants. Adults (age 18–69 years) with SSc meeting study entry criteria were eligible for randomization (4). Participants secured approval from health insurers for study treatments (including appeal of coverage denials), provided informed consent for study screening and treatment, and were enrolled from July 2005 through September 2011.

Study design. Seventy-five participants were randomized to either myeloablative autologous HSCT or 12 monthly infusions of CYC. Details of the study design, mobilization, and selection of CD34+ cells, preparative conditioning, autologous transplantation, CYC administration, and posttreatment care have been previously reported (4).

A data and safety monitoring board provided oversight. Site institutional review boards approved the protocol. Rho, Inc. (Durham, NC) held and analyzed the data. Members of the Steering Committee designed the trial and attest to the fidelity of the data and analyses. The first author wrote the initial draft. All coauthors reviewed the manuscript and agreed to publication.

Evaluations and end points. Assessments were performed at baseline, weeks 4 and 12, and months 8, 14, 20, 26, 32, 44, 48, 54, 60, 66, and 72 and included the following: FVC and DLco (both % of predicted), HAQ DI score (range 0 [no impairment] to 3 [completely impaired]), MRSS (range 0 [normal] to 51 [severe]), and physical and mental composite scores from the SF-36 (normalized to the US population [1998] with a mean ± SD of 50 ± 10; lower scores indicating poorer quality of life). The MRSS was not assessed at weeks 4 and 12 and months 20, 32, 44, 60, and 72. Times are in relation to the date of randomization. For FVC, DLco, and SF-36 mental and physical composites, lower scores are poorer outcomes. For HAQ DI and MRSS, lower scorers are better outcomes. SSc intrinsic molecular subset assignments (inflammatory, fibroproliferative, and normal-like) were derived from gene expression analysis of baseline peripheral blood samples using a machine learning classifier, as previously described (11,12,21).

Longitudinal analyses and statistical models. The study design called for at least 54 months of follow-up, but data collection ceased early for individuals who withdrew, experienced organ failure, or died. Because organ failure and death contributed to early loss of subjects, which was differentially distributed between treatment arms, methods herein account for data that

are not "missing completely at random." We considered 2 possible assumptions about the nature of the relationship between unobserved responses and time on study.

First, we assumed that unobserved responses were not impacted by time on study. For example, if a participant experienced pulmonary failure and stopped attending clinic visits, then one plausible assumption is that unobserved future FVC or DLco responses would continue on a trajectory consistent with past observations. This is consistent with the "missing at random" assumption, which implies that the observed response data are sufficient for estimating the overall expected response trajectory. Under the "missing at random" assumption, mixed-effects regression models were used for interpolation of longitudinal trends. In all models described below, smoking status (never versus past/present), prestudy use of CYC (yes/no), and sex (male/female) were included as covariates.

For FVC and DLco, mixed-effects regression models were fit as splines for data interpolation with separate intercepts and slopes for each treatment (22). For HSCT, the spline included pivot points at 3 and 14 months posttransplant. At 3 months, pulmonary function was expected to fall after total body irradiation. At 14 months, recovery was expected to be complete. For CYC, postbaseline pulmonary tests were not completed before 8 months; the spline included a single pivot at 14 months (2 months after CYC treatments concluded). Pulmonary function testing at SCOT centers was included as a time-varying covariate. Subject-level effects for intercept, initial slope, and post month-14 slope were included as random effects assuming separate banded (2) unstructured covariance structures for each treatment. The mixed-effects regression models for quality of life outcomes are analogous to the FVC and DLco models, with several caveats. Splines for both arms had a single pivot point at month 14. The pulmonary function test site was not included as a time-varying covariate. For MRSS, because the rate of decline decreases over time, a negative exponential decay model was used to interpolate longitudinal trends. The nonlinear mixed model included separate fixed intercepts and decay constants for each treatment arm. Random subject-level effects for intercept and decay constant were included, assuming separate unstructured covariance structures for each treatment. Mixed-effects regression models to explore intrinsic gene subsets included separate intercepts and slopes (or decay rates) for each treatment-by-intrinsic-subset group.

Second, we considered the possibility that the rate of dropout over time and unobserved responses were associated, which implies that data are missing not at random and that the expected response trajectory cannot be reliably predicted from observed response data alone. One could reasonably envision a scenario in which an individual's time on study is correlated with baseline values and/or the change in response over time (slopes or decay rate). To explore this possibility, we modeled the longitudinal response and hazard rates for time on study simultaneously using shared parameter models. To fit the shared parameter model, the mixed-effects regression model described above served as the longitudinal component. The time-on-study component was modeled as a piecewise exponential survival model with separate baseline log-failure rates for each treatment arm over each of 2 intervals: 0–24 months, and 24 months to end of participation. The random subject-level effects for intercept, initial slope during the treatment period, and post month-14 slope (or decay rate) from the longitudinal component function served as fixed covariates in the time-on-study survival model.

Both mixed and joint longitudinal/survival models were prespecified secondary analyses aimed at evaluating long-term trends under different assumptions about missing data. Results from mixed-effects regression models are the primary focus for inference, and those for the shared parameter models are presented as sensitivity analyses. Missing data analyses with mixed-effects regression and shared parameter approaches have been described (10).

P values are from model-derived *t*-tests. Key time points were selected for comparison: baseline, month 14 (CYC treatment is completed, and recovery from transplant is expected), and month 54 (indicative of long-term benefit). These secondary analyses explore relationships in order to better understand treatment differences. *P* values are not adjusted for multiple comparisons, but those ≤0.05 are highlighted as notable findings. Analyses used SAS, version 9.4.

RESULTS

Study participants. In order to describe treatment-specific longitudinal trends, the analysis sample was limited to 67 of 75 SCOT participants (33 of 36 HSCT, 34 of 39 CYC) (4) who actually received a transplant or completed ≥9 CYC doses. The 3 excluded from the HSCT arm include 2 who became ineligible for transplant and 1 who died prior to initiating the procedure. The 5 excluded from the CYC arm include 3 who withdrew consent (2 with no doses, 1 with 2 CYC doses) and 2 who died prior to completing dosing (1 each with 2 and 5 CYC doses). Baseline characteristics for this analysis sample are comparable to those reported for all SCOT participants (4) and reflect severe scleroderma: mean baseline MRSS = 29, DLco = 53% of predicted, and 99% with pulmonary involvement (see Supplementary Table 1, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24785). Both arms had similar characteristics except that the CYC arm included more female participants, never smokers, and prior use of CYC.

Longitudinal trends for clinical disease measures and quality of life outcomes. Figure 1 depicts individual trajectories for FVC, which generally differ between arms. For HSCT, the decline in FVC at 3 months with recovery by 14 months is expected following irradiation. For CYC, observed FVC values

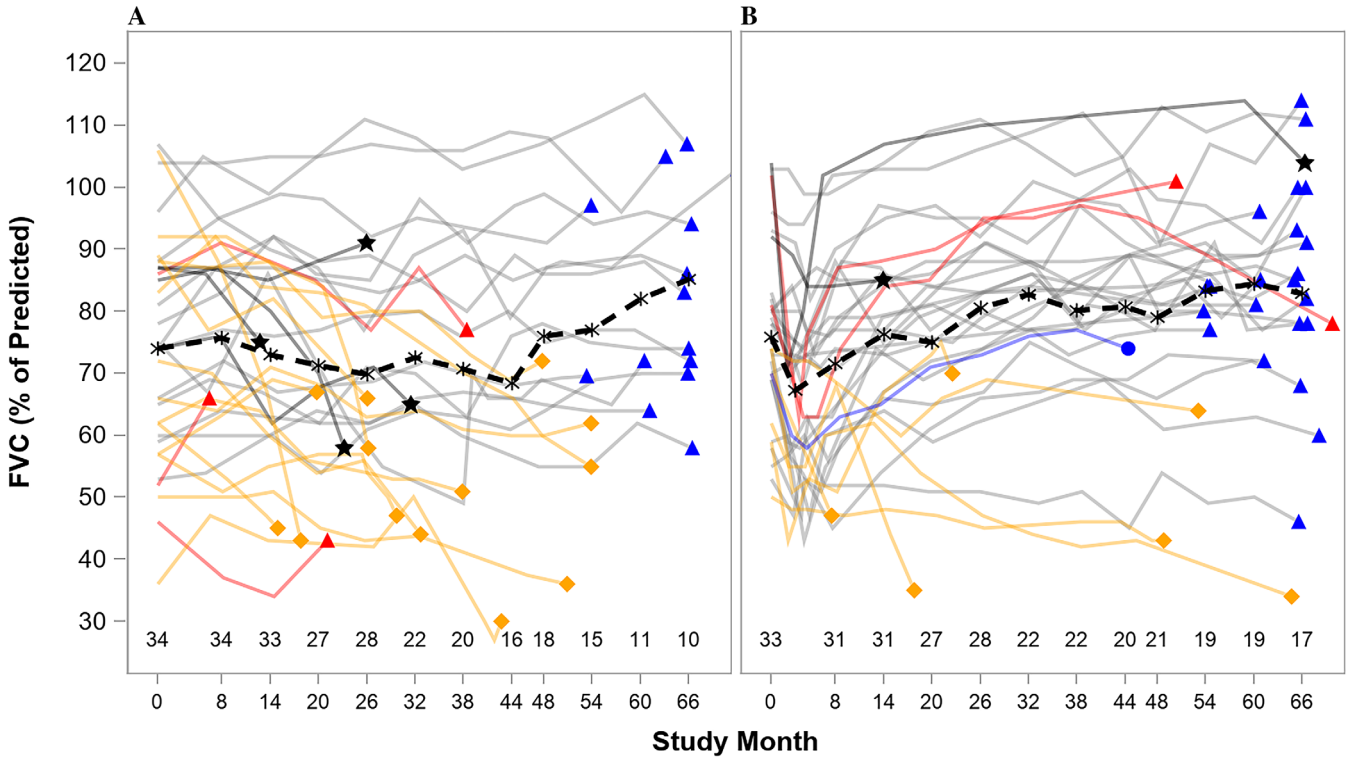

Figure 1. Forced vital capacity (FVC) trajectories with reasons for termination in the cyclophosphamide (**A**) and transplant (**B**) arms. Each line connects FVC values (% predicted) over time for an individual participant. Line colors and symbols at the last assessment indicate reasons for termination: completed study per protocol (blue triangle), death (black star), early withdrawal (red triangle), early last FVC (blue circle), organ failure (gold diamond). The broken black line connects the means (black asterisks) for available data clustered within the following time intervals in months (midpoint [range]): 0, 8 (5–11), 14 (11–17), 20 (17–23), 26 (23–29), 32 (29–35), 38 (35–41), 44 (41–47), 48 (47–51), 54 (51–57), 60 (57–63), and 66 (63 and above). Numbers of subjects available for each interval are given above the x-axis.

display random variation about relatively linear trajectories. Importantly, premature cessation of data collection occurred more often in CYC recipients. Eighteen (53%) CYC participants had their last FVC assessment before 54 months compared to 7 (21%) HSCT recipients. In both arms, the observed means for FVC increased over time (Figure 1 [broken line]). However, per the SCOT protocol, follow-up pulmonary assessments ceased once a participant experienced respiratory failure, defined by DLCO or FVC criteria. Respiratory failure accounts for the majority of early terminations (13 CYC, 5 HSCT). As such, due to early loss of those with poor FVCs, the means for survivors increase over time, giving potentially favorably biased representations of expected FVC overall response trajectories, particularly for the CYC population, where attrition was greater.

Within each treatment population, we assume that individual trajectories vary about an expected overall response trajectory for the entire population. To compare the expected overall response trajectories for HSCT and CYC, we are compelled to make assumptions about the unknown nature of missing response data. Assuming that individuals lost to follow-up early would have continued on their same trajectories regardless of time on study, observed responses for individuals are sufficient for estimating the expected overall response trajectory. This is the missing-at-random assumption. Although validity of the assumption cannot be confirmed, it is consistent with the fact that respiratory data are missing by design for those who experienced respiratory failure. If valid, then a mixed-effects regression model provides unbiased estimates of longitudinal trends.

With the mixed-effects regression approach, the expected overall FVC trajectories differ between the treatment arms (Figure 2A [solid lines] and Table 1). At time 0, expected values are 75.9 for HSCT and 74.6 for CYC ($P = 0.754$). After the initial fall due to irradiation in the HSCT arm, FVC recovers by month 14 with an expected value of 77.4 compared to 72.8 for CYC ($P = 0.299$). After month 14, improvement in FVC for HSCT was sustained, with an expected increase of 0.77 percentage points/year compared to an average fall of –3.70 percentage points/year for CYC ($P = 0.004$). By month 54, there are notable differences between arms (expected values: 79.9 HSCT versus 60.4 CYC; $P = 0.005$). Shared parameter models (broken lines) are consistent with the mixed-effects regression models (solid lines). The available observed means at each assessment point (clustered as in Figure 1) for both treatments show how failure to account for missing data due to early loss of participants can impact estimated trends.

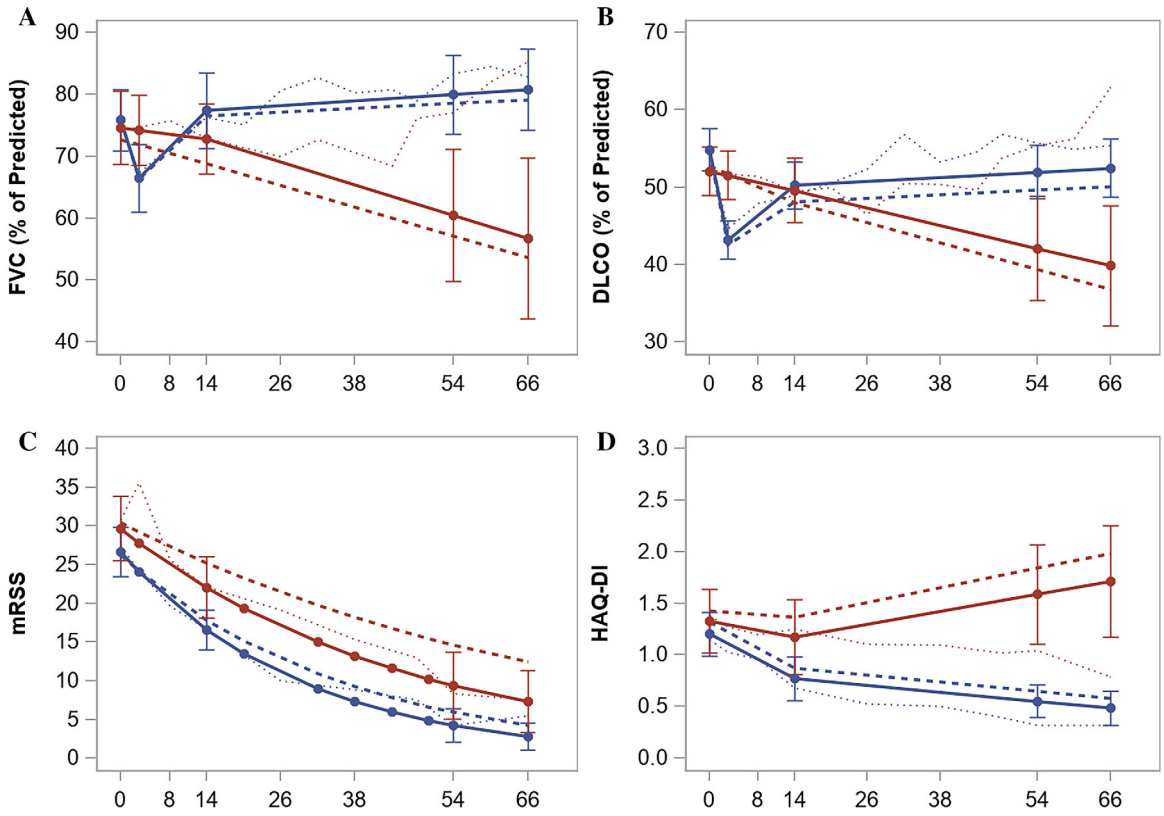

Figure 2. Estimated longitudinal trends for mixed-effects regression and shared parameter models for forced vital capacity (FVC) (**A**), diffusing capacity for carbon monoxide (DLco) (**B**), modified Rodnan skin thickness score (mRSS) (**C**), and Health Assessment Questionnaire disability index (HAQ-DI) (**D**). Estimates for the hematopoietic cell transplant and cyclophosphamide arms are presented in blue and red, respectively. Solid lines show model-based estimates for mixed-effects regression models; vertical lines show 95% confidence intervals at select time points. Broken lines show model-based means for shared parameter models. Broken lines connect means for available data clustered as per Figure 1.

Mixed-effects regression findings for DLco are consistent with FVC (Figure 2B [solid lines] and Table 1). Expected time 0 values are 54.8 for HSCT and 52.1 for CYC (P = 0.212). By month 14, expected values were 50.2 for HSCT compared to 49.6 for CYC (P = 0.813). After month 14, DLco improved modestly for HSCT, with an expected increase of 0.52 percentage points/year compared to an average fall of –2.25 percentage points/year for CYC ($P \leq 0.001$). As with FVC, the benefit of transplant is apparent by month 54 (expected values: 52.0 HSCT, 42.1 CYC; P = 0.016).

In both arms, MRSS declined (improved) exponentially over time (Figure 2C; see Supplementary Table 2, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24785). For the exponential mixed-effects regression model, time 0 values were 26.7 for HSCT and 29.6 for CYC (P = 0.259). Notable differences favoring HSCT were observed at time points starting at month 14: MRSS fell faster for transplant, with a decay rate of 0.41 per year for HSCT compared to 0.26 per year for CYC (P = 0.048).

For both the HAQ DI and SF-36 physical composite scores, the mixed-effects regression approach shows that quality of life improves for HSCT and declines for CYC, with notable differences by month 54. Expected values at month 54 for HAQ DI are 0.55 and 1.59 for HSCT and CYC, respectively ($P < 0.001$) (Figure 2D; see Supplementary Table 3, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24785). For the SF-36 physical composite score, expected values at month 54 are 45.1 HSCT and 32.6 CYC ($P < 0.001$; see Supplementary Figure 1 and Supplementary Table 4, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24785). In contrast, trends for the SF-36 mental composite scores do not differ between arms, and expected values over all time points are consistent with the normalizing population (i.e., 1998 US population mean ± SD 50 ± 10; see Supplementary Figure 1 and Supplementary Table 5, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24785).

If the expected overall response trajectory depends on both the response data and time on study, then data for individuals lost to follow-up early are missing not at random. Results from shared parameter models fit under this missing-not-at-random assumption confirm the findings from the mixed-effects regression analyses for FVC, DLco, MRSS, HAQ DI, and the SF-36 composite scores (Figure 2 and Table 1; see Supplementary Figure 1 and Supplementary Tables 2–5, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24785). In some cases, shared parameter

Table 1. Longitudinal trends for forced vital capacity (FVC) % predicted and diffusing capacity for carbon monoxide (DLco) % predicted: mixed-effects regression and shared parameter models*

	Mixed effects regression, expected value (SE)	Shared parameter, expected value (SE)
FVC		
Time 0		
Transplant	75.9 (2.51)	75.9 (2.67)
Cyclophosphamide	74.6 (2.98)	72.7 (3.11)
P	0.754	0.451
Month 14		
Transplant	77.4 (3.12)	76.5 (3.54)
Cyclophosphamide	72.8 (2.89)	68.8 (3.41)
P	0.299	0.139
Month 54		
Transplant	79.9 (3.21)	78.5 (3.72)
Cyclophosphamide	60.4 (5.42)	57.1 (8.64)
P	0.005†	0.025†
Slope after month 14 (per year)		
Transplant	0.77 (0.39)	0.61 (0.43)
Cyclophosphamide	–3.70 (1.38)	–3.50 (2.50)
P	0.004†	0.109
DLco		
Time 0		
Transplant	54.8 (1.38)	54.5 (1.36)
Cyclophosphamide	52.1 (1.62)	52.8 (1.58)
P	0.212	0.437
Month 14		
Transplant	50.2 (1.56)	48.1 (1.85)
Cyclophosphamide	49.6 (2.12)	48.0 (2.78)
P	0.813	0.964
Month 54		
Transplant	52.0 (1.74)	49.6 (2.31)
Cyclophosphamide	42.1 (3.42)	39.4 (7.15)
P	0.016†	0.179
Slope after month 14 (per year)		
Transplant	0.52 (0.34)	0.45 (0.40)
Cyclophosphamide	–2.25 (0.64)	–2.57 (1.91)
P	<0.001†	0.125

* Times are in relation to the date of randomization.
† Significant.

trajectories point toward worse outcomes. For example, the expected overall FVC trajectories from mixed-effects regression and shared parameter models track closely for HSCT recipients (Figure 2A), suggesting that time on study has little impact. In contrast, for CYC recipients, the expected overall FVC trajectory derived from the shared parameter model is shifted toward lower FVCs relative to the mixed-effects regression model.

For the SF-36 mental composite score, a convergent shared parameter model could not be found. However, the observed means track well with the expected overall response trajectories estimated from the mixed-effects regression model, suggesting that missing data have little impact on this end point (see Supplementary Figure 1, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24785).

Baseline intrinsic molecular subsets and longitudinal trends. Sixty-two (93%) of the 67 per-protocol study participants were categorized into intrinsic subsets: 20 inflammatory (8 HSCT, 12 CYC), 22 normal-like (10 HSCT, 12 CYC), and 20 fibroproliferative (11 HSCT, 9 CYC).

For FVC and DLco (Figures 3 and 4 and Table 2), trends for HSCT and CYC did not differ notably for the normal-like subset. For both inflammatory and fibroproliferative subsets, trends favored HSCT. At month 14, differences between arms are not notable. In the inflammatory and fibroproliferative subsets after month 14, FVC worsened for CYC (–4.61 and –5.62 percentage points/year [slope], respectively) and improved in the HSCT arm (0.47 and 1.78 percentage points/year, respectively; $P = 0.054$ inflammatory, $P = 0.012$ fibroproliferative). Similarly, for inflammatory and fibroproliferative, DLco worsened for CYC (–2.56 and –3.10 percentage points/year, respectively) and improved for HSCT (1.51 and 0.47 percentage points/year, respectively; $P = 0.003$ inflammatory, $P = 0.016$ fibroproliferative).

For MRSS (see Supplementary Figure 2 and Supplementary Table 6, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24785), exponential

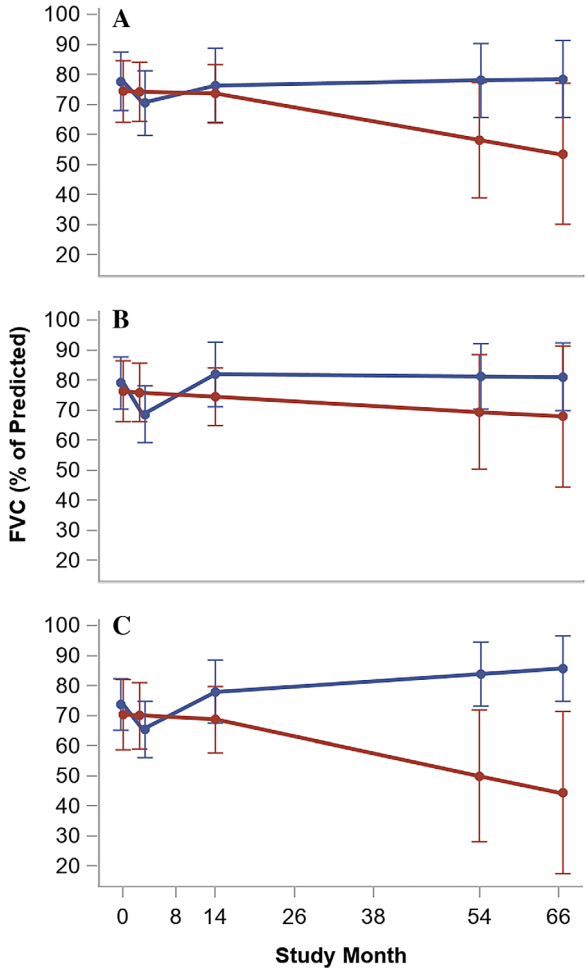

Figure 3. Trends from mixed-effects regression models for intrinsic molecular subsets: inflammatory (**A**), normal-like (**B**), and fibroproliferative (**C**). Forced vital capacity (FVC) estimates for the transplant and cyclophosphamide arms are presented in blue and red, respectively. Solid lines show model-based estimates for mixed-effects regression models; vertical lines show 95% confidence intervals at select time points. Mixed-effects regression models are as described for Figure 2A with the addition of separate intercepts and slopes (or decay rates) for each treatment-by-intrinsic-subset group.

decay rates favoring HSCT differ notably between arms for only the fibroproliferative subset (decay rates/year = 0.522 HSCT, 0.238 CYC; P = 0.011). For HAQ DI (see Supplementary Figure 2 and Supplementary Table 7, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24785), trends favoring HSCT are similar for all subsets. Treatment arms did not differ notably at month 14. After month 14, HAQ DI score worsened for CYC (points/year change = 0.191 inflammatory, 0.123 normal-like, and 0.069 fibroproliferative) and improved for HSCT (points/year change = –0.036 inflammatory, –0.066 normal-like, and –0.090 fibroproliferative). Expected values differed notably for CYC and HSCT at month 54 for all subsets (P = 0.016 inflammatory, 0.017 normal-like, 0.017 fibroproliferative).

For the SF-36 physical composite score (see Supplementary Figure 3 and Supplementary Table 8, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24785), estimates favored HSCT for all 3 subsets but are evident earlier in the fibroproliferative subset, where differences in expected values are detected at month 14 (45.2 HSCT, 26.3 CYC; $P < 0.001$) and retained through month 54 (47.7 HSCT, 28.0 CYC; $P < 0.001$). For the inflammatory subset, differences in expected values are notable at month 54 (inflammatory: 44.6 HSCT, 32.8 CYC; P = 0.026). Trends for the SF-36 mental composite score did not differ notably between treatment groups for the normal-like or fibroproliferative subsets (see Supplementary Figure 3 and Supplementary Table 9, available on the *Arthritis*

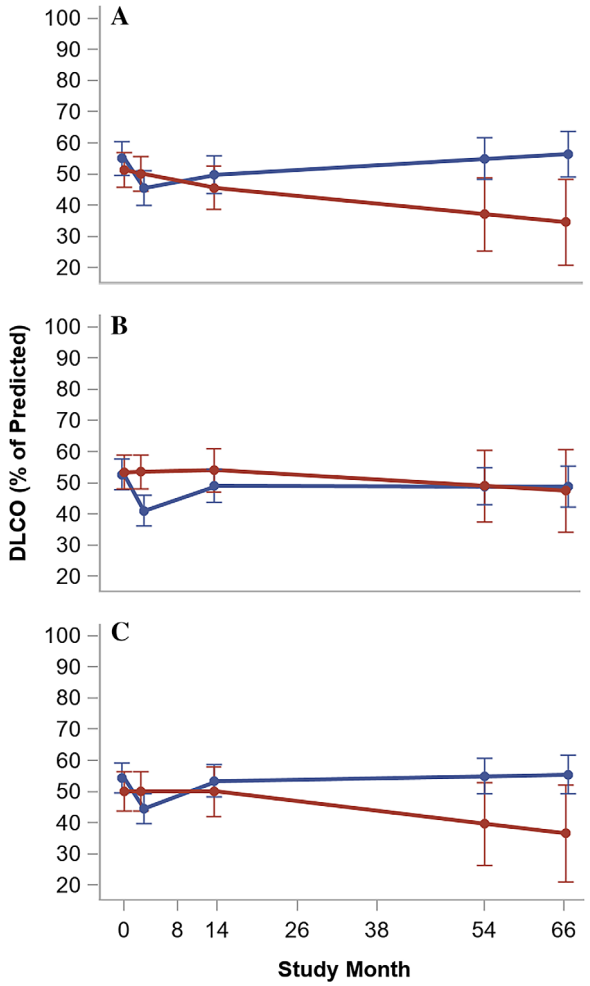

Figure 4. Trends from mixed-effects regression models for intrinsic molecular subsets: inflammatory (**A**), normal-like (**B**), and fibroproliferative (**C**). Diffusing capacity for carbon monoxide (DL_{CO}) estimates for the transplant and cyclophosphamide arms are presented in blue and red, respectively. Solid lines show model-based estimates for mixed-effects regression models; vertical lines show 95% confidence intervals at select time points. Mixed-effects regression models are as described for Figure 2B with the addition of separate intercepts and slopes (or decay rates) for each treatment-by-intrinsic-subset group.

Table 2. Longitudinal trends for forced vital capacity (FVC)% predicted and diffusing capacity for carbon monoxide (DLco)% predicted by baseline intrinsic molecular subsets

	Inflammatory, expected value (SE)	Normal-like, expected value (SE)	Fibroproliferative expected value (SE)	*P**
FVC				
Time 0				
Transplant	77.8 (4.99)	79.2 (4.39)	73.8 (4.36)	0.666
Cyclophosphamide	74.5 (5.23)	76.4 (5.15)	70.5 (5.92)	0.750
P†	0.656	0.679	0.650	
Month 14				
Transplant	76.5 (6.23)	82.0 (5.50)	78.1 (5.37)	0.784
Cyclophosphamide	73.7 (4.98)	74.6 (4.90)	68.8 (5.60)	0.714
P†	0.729	0.321	0.234	
Month 54				
Transplant	78.1 (6.27)	81.4 (5.55)	84.0 (5.40)	0.770
Cyclophosphamide	58.3 (9.75)	69.5 (9.70)	50.1 (11.14)	0.412
P†	0.092	0.291	0.007‡	
Slope after month 14 (per year)				
Transplant	0.47 (0.78)	−0.19 (0.66)	1.78 (0.61)	0.085
Cyclophosphamide	−4.61 (2.51)	−1.53 (2.46)	−5.62 (2.86)	0.508
P†	0.054	0.597	0.012‡	
DLco				
Time 0				
Transplant	55.1 (2.78)	52.8 (2.47)	54.5 (2.43)	0.802
Cyclophosphamide	51.4 (2.83	53.5 (2.80)	50.1 (3.23)	0.721
P†	0.355	0.846	0.286	
Month 14				
Transplant	50.0 (3.08)	49.2 (2.71)	53.4 (2.63)	0.482
Cyclophosphamide	45.7 (3.57)	54.1 (3.56)	50.0 (4.05)	0.254
P†	0.372	0.276	0.481	
Month 54				
Transplant	55.0 (3.40)	48.9 (3.04)	55.0 (2.90)	0.268
Cyclophosphamide	37.2 (6.01)	49.0 (5.86)	39.7 (6.78)	0.336
P†	0.011‡	0.989	0.038‡	
Slope after month 14 (per year)				
Transplant	1.51 (0.64)	−0.07 (0.55)	0.47 (0.51)	0.172
Cyclophosphamide	−2.56 (1.18)	−1.51 (1.06)	−3.10 (1.28)	0.607
P†	0.003‡	0.226	0.016‡	

* This *P* value is for a comparison of the 3 intrinsic subsets within the treatment arm.
† This *P* value is for a comparison of transplant and cyclophosphamide.
‡ Significant.

Care & Research website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24785). In the inflammatory subset, the trend after month 14 is stable for transplant and worsens for CYC, resulting in differences at month 54 (inflammatory expected value = 54.6 HSCT, 42.5 CYC; $P = 0.019$).

DISCUSSION

Failure to account for early loss of participant data may distort estimates of clinical trends over time. To account for this potential distortion, we applied statistical techniques to account for missing data under 2 different assumption: missing at random and missing not at random. With both approaches, CYC recipients had clinical measures worsen over time compared to HSCT recipients. Longitudinal trends for inflammatory and fibroproliferative intrinsic molecular subsets showed similar responses over time after accounting for dropouts.

These results demonstrating the superiority for transplant over CYC for clinical responses and quality of life over the long-term provide a clinical context that strengthens the previously reported results of the GRCS analysis at 54 months and support the validity of this measure (4). Results for FVC, HAQ DI, and MRSS also establish their value as hierarchical components of the GRCS for the study of SSc with organ involvement. The value of the GRCS for other subpopulations of scleroderma requires further investigation.

Missing data present challenges for trials in rheumatic and other diseases (23–31). Among randomized treatment studies of rheumatoid arthritis with composite outcomes, >30% rates of missing data were noted in 9 of 51 (17%) trials reported between 2008 and 2013 (32). Statistical methodologies to approach such bias have been developed (6–8,10,33–35). In the ASSIST and ASTIS trials, which also compared HSCT to CYC in SSc, longitudinal results for FVC and MRSS were based on subsets of

individuals with data at 12–24 months after randomization (36,37). This complete-case approach was not a valid option for examining long-term trends in the SCOT trial because loss of participants prior to 54 months was at least partially due to disease manifestations. Trends for this responder subset are not representative of those randomized.

Our first objective was to estimate and compare anticipated long-term trends while adjusting for possibly informative censoring due to premature loss of participants. We considered 2 possible assumptions about the nature of the relationship between unobserved responses and time on study. The underlying assumption for the mixed-effects regression models is that unobserved future responses would continue on a trajectory consistent with past observations (i.e., missing at random). For the shared parameter models, the assumption is that the rate of dropout over time and unobserved responses are associated (i.e., missing not at random). We found that while estimates from the mixed-effects regression and shared parameter models differ slightly numerically, findings consistently show superiority of transplant for all outcomes investigated except for the SF-36 mental composite score, where impairment appeared minimal in this study population.

The second objective was to explore how longitudinal trends might differ depending on baseline intrinsic molecular subset assignment. Franks et al found that the event-free survival advantage of HSCT over CYC in SCOT was most pronounced for the fibroproliferative subset, less definitive for the inflammatory subset, and absent in the normal-like subset (21). In the present analyses, HSCT recipients in the fibroproliferative subset had superior long-term trends compared to CYC on all clinical and quality of life responses except the SF-36 mental composite, including the most rapid improvements in MRSS and SF-36 physical composite score. Transplant recipients in the inflammatory subset had superior long-term trends compared to CYC on all clinical and quality of life responses except the MRSS. In the normal-like subset, although HSCT recipients often had numerically superior trends relative to CYC, findings were only notably different for HAQ DI score. Overall, these findings suggest that transplant could be most effectively applied for individuals with fibroproliferative or inflammatory signatures. The inflammatory and fibroproliferative subsets may represent active states of SSc, so improved outcomes with HSCT for these subsets may be attributable to termination of the underlying ongoing immune processes. For the fibroproliferative subset, results from prior studies on immune modulators such as mycophenolate mofetil (15,38) or abatacept (18) that have failed to demonstrate robust responses are consistent with our findings showing little benefit in the CYC arm across multiple end points. The normal-like subset may represent individuals who no longer have active immune processes but still have tissue damage, which may not be amenable to therapy with immunologic modulators. In the present study, however, the negative findings in the normal-like group are uninformative. Given the small numbers and exploratory nature of these analyses, we cannot conclude that transplant has no benefit in the normal-like group. These possibilities warrant further study.

Our study has limitations. Because data are unavailable, the missing data assumptions cannot be verified and hence should be acknowledged when evaluating model results. Further, the interpretation of model estimates in the context of these assumptions requires clarification. Typically, model-based point estimates from a linear regression model can be interpreted as population estimates, but not in the present situation. Because responses do not exist for individuals who died prior to month 54, the 54-month estimate cannot represent a mean for the entire population. Rather, model-based estimates presented here are expected response trajectories and expected values at given time points for an individual, conditional on survival.

In conclusion, failure to account for early loss of participant data may distort estimates of clinical trends over time. Using 2 statistical models for missing data approaches, long-term clinical measures validate the superiority of HSCT over CYC for treatment of severe scleroderma. Such approaches accounting for longitudinal bias due to missing trial data may more accurately guide future care of patients with rheumatic diseases.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be submitted for publication. Dr. Sullivan had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Keyes-Elstein, Pinckney, Goldmuntz, Welch, Franks, Martyanov, Wood, Crofford, Mayes, McSweeney, Nash, Georges, Csuka, Simms, Furst, Khanna, St.Clair, Whitfield, Sullivan.

Acquisition of data. Keyes-Elstein, Pinckney, Goldmuntz, Welch, Franks, Martyanov, Wood, Crofford, Mayes, McSweeney, Nash, Georges, Csuka, Simms, Furst, Khanna, St.Clair, Whitfield, Sullivan.

Analysis and interpretation of data. Keyes-Elstein, Pinckney, Goldmuntz, Welch, Whitfield, Sullivan.

ADDITIONAL DISCLOSURES

Authors Keyes-Elstein and Pinckney are employees of Rho.

REFERENCES

1. Mayes MD, Lacey JV Jr, Beebe-Dimmer J, et al. Prevalence, incidence, survival, and disease characteristics of systemic sclerosis in a large US population. Arthritis Rheum 2003;48:2246–55.
2. Elhai M, Meune C, Avouac J, et al. Trends in mortality in patients with systemic sclerosis over 40 years: a systematic review and meta-analysis of cohort studies. Rheumatology (Oxford) 2012;51:1017–26.
3. Nihtyanova SI, Schreiber BE, Ong VH, et al. Prediction of pulmonary complications and long-term survival in systemic sclerosis. Arthritis Rheumatol 2014;66:1625–35.
4. Sullivan KM, Goldmuntz EA, Keyes-Elstein L, et al. Myeloablative autologous stem-cell transplantation for severe scleroderma. N Engl J Med 2018;378:35–47.

5. Assassi S, Wang X, Chen G, et al. Myeloablation followed by autologous stem cell transplantation normalizes systemic sclerosis molecular signatures. Ann Rheum Dis 2019;78:1371–8.
6. Little RJ, D'Agostino R, Cohen ML, et al. The prevention and treatment of missing data in clinical trials. N Engl J Med 2012;367: 1355–60.
7. Bell ML, Fiero M, Horton NJ, et al. Handling missing data in RCTs;a review of the top medical journals. BMC Med Res Methodol 2014; 14:118–25.
8. Ibrahim JG, Chu H, Chen MH, et al. Missing data in clinical studies: issues and methods. J Clin Oncol 2012;30:3297–303.
9. Zhang Y, Florez ID, Lozano LE, et al. A systematic survey on reporting and methods for handling missing participant data for continuous outcomes in randomized controlled trials. J Clin Epidemiol 2017;88: 57–66.
10. Vonesh EF. Generalized linear and nonlinear models for correlated data: theory and applications using SAS®. Cary (NC): SAS Institute Inc; 2012.
11. Whitfield ML, Finlay DR, Murray JL, et al. Systemic and cell type-specific gene expressions in scleroderma skin. PNAS 2003;100: 12319–24.
12. Franks JM, Martyanov V, Cai G, et al. A machine learning classifier for assigning individual patients with systemic sclerosis to intrinsic molecular subsets. Arthritis Rheumatol 2019;71:1701–10.
13. Milano A, Pendergrass SA, Sargent JL, et al. Molecular subsets in the gene expression signatures of scleroderma skin. PLoS One 2008;3: e2696.
14. Pendergrass SA, Lemaire R, Francis IP, et al. Intrinsic gene expression subsets of diffuse cutaneous systemic sclerosis are stable in serial skin biopsies. J Invest Dermatol 2012;132:1363–73.
15. Hinchcliff M, Huang CC, Wood TA, et al. Molecular signatures in skin associated with clinical improvement during mycophenolate treatment in systemic sclerosis. J Invest Dermatol 2013;133:1979–89.
16. Mahoney JM, Taroni J, Martyanov V, et al. Systems level analysis of systemic sclerosis shows a network of immune and profibrotic pathways connected with genetic polymorphisms. PLoS Comput Biol 2015;11:e1004005.
17. Skaug B, Khanna D, Swindell WR, et al. Global skin gene expression analysis of early diffuse cutaneous systemic sclerosis shows a prominent innate and adaptive inflammatory profile. Ann Rheum Dis 2020; 79:379–86.
18. Khanna D, Spino C, Johnson S, et al. Abatacept in early diffuse cutaneous systemic sclerosis: results of a phase II investigator-initiated, multicenter, double-blind, randomized, placebo-controlled trial. Arthritis Rheumatol 2020;72:125–36.
19. Taroni JN, Martyanov V, Huang CC, et al. Molecular characterization of systemic sclerosis esophageal pathology identifies inflammatory and proliferative signatures. Arthritis Res Ther 2015;17:194.
20. Taroni JN, Greene CS, Martyanov V, et al. A novel multi-network approach reveals tissue-specific cellular modulators of fibrosis in systemic sclerosis. Genome Med 2017;9:27.
21. Franks JM, Martyanov V, Wang Y, et al. Machine learning predicts stem cell transplant response in severe scleroderma. Ann Rheum Dis 2020;79:1608–15.
22. Gauthier J, Wu QV, Gooley TA. Cubic splines to model relationships between continuous variables and outcomes: a guide for clinicians. Bone Marrow Transplant 2020;55:675–80.
23. Baron G, Boutron I, Giraudeau B, et al. Violation of intent-to-treat principle and rate of missing data in superiority trials assessing structural outcomes in rheumatic diseases. Arthritis Rheumatol 2005;54: 1858–65.
24. Wong WK, Boscardin WJ, Postlethwaite AE, et al. Handling missing data issues in clinical trials for rheumatic diseases. Contemp Clin Trials 2011;1–9.
25. Fleming TR. Addressing missing data in clinical trials. Ann Intern Med 2011;154:1113–17.
26. Akl EA, Briel M, You JJ, et al. Potential impact on estimated treatment effects of information lost to follow-up in randomized controlled trials (LOST-IT): systematic review. Brit Med J 2012;344:e2809.
27. Fielding S, Ogbuagu A, Sivasubramaniam S, et al. Reporting and dealing with missing quality of life data in RCTs: has the picture changed in the last decade? Qual Life Res 2016;25:2977–83.
28. Symmons DP. Epidemiology research in rheumatology-progress and pitfalls. Nat Rev Rheumatol 2015;11:631–8.
29. Kim M, Merrill JT, Wang C, et al. SLE clinical trials: impact of missing data on estimating treatment effects. Lupus Sci Med 2019;6: e000348.
30. Budhiraja P, Kalot M, El Alayli E, et al. Reporting and handling of missing participant data in systematic reviews of kidney transplant studies. Transplantation 2021;105:1708–17.
31. Kahale LA, Khamis AM, Diab B, et al. Potential impact of missing outcome data on treatment effects in systematic reviews: imputation study. Brit Med J 2020;370:m2898.
32. Ibrahim F, Tom BD, Scott DL, et al. A systematic review of randomized controlled trials in rheumatoid arthritis: the reporting and handling of missing data in composite outcomes. Trials 2016;17:272–80.
33. Bell ML, Kenward MG, Fairclough DL, et al. Differential dropout and bias in RCs: when it matters and when it may not. Brit Med J 2013; 346:e8668.
34. Powney M, Williamson P, Kirkham J, et al. A review of the handling of missing longitudinal outcome data in clinical trials. Trials 2014;15:237.
35. Zhang Y, Alyass A, Vanniyasingam T, et al. A systematic survey of the methods literature on the reporting quality and optimal methods of handling participants with missing outcome data for continuous outcomes in randomized controlled trials. J Clin Epidemiol 2017;889: 67–80.
36. Burt RK, Shah SJ, Dill K, et al. Autologous non-myeloablative haemopoietic stem-cell transplantation compared with pulse cyclophosphamide once per month for systemic sclerosis (ASSIST): an open-label, randomized phase 2 trial. Lancet 2011;378:498–506.
37. Van Laar JM, Farge D, Sont JK, et al. Autologous hematopoietic stem cell transplantation vs intravenous pulse cyclophosphamide in diffuse cutaneous systemic sclerosis: a randomized clinical trial. JAMA 2014; 311:2490–8.
38. Hinchcliff M, Toledo DM, Taroni JN, et al. Mycophenolate mofetil treatment of systemic sclerosis reduces myeloid cell numbers and attenuates the inflammatory gene signature in skin. J Invest Dermatol 2018; 138:1301–10.