SUPPLEMENTARY MATERIALS 1: Methods

2023 American College of Rheumatology (ACR)/American College of Chest Physicians (CHEST) Guideline for the Treatment of Interstitial Lung Disease in People with Systemic Autoimmune Rheumatic Diseases

Methodology Overview

This guideline was developed following the American College of Rheumatology (ACR) guideline development process (www.rheumatology.org/Portals/0/Files/ACR%20Guideline%20Manual_Appendices_updated%202015.pdf). This process includes the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) methodology (www.gradeworkinggroup.org) (1-4).

Teams Involved

Four groups with distinct responsibilities were involved in the development of these guidelines: the Core Team, Literature Review Team, Patient Panel, and Voting Panel.

A Core Leadership Team (11 members) met weekly to supervise the project and was responsible for confirming the scope and developing clinical (Patient/Intervention/Comparator/Outcomes – PICO) questions (see Supplementary Materials 2), coordinating with the Literature Review Team, overseeing the voting process, and drafting the manuscript. The Core Team, together with the Literature Review Team, was comprised of individuals with content and methodological expertise, and included a GRADE methodologist who advised on the process of developing and presenting the evidence and provided input on the quality assessment of evidence and summary of findings (SoF) tables (provided in Supplementary Materials 3).

The Literature Review Team (22 members) conducted a systematic search with the assistance of an experienced medical librarian. This group reviewed articles for relevance, assessed study quality, extracted data, computed pooled estimates of outcomes, graded the quality of evidence, generated an evidence summary for each PICO, and compiled an evidence report.

A Patient Panel was convened to discuss patient values and preferences related to outcomes, evidence, and drafted recommendation statements. The ACR solicited volunteers for the Patient Panel, collecting details regarding interstitial lung disease (ILD) disease experience, experience with therapies under consideration, and potential conflicts of interest. The Core Team reviewed the applications to select members for the Patient Panel including 3 patients to participate on the Voting Panel. The

Voting Panel used the input from the Patient Panel meeting to help guide their votes in balancing tradeoffs between the harms and benefits of the alternative management strategies.

The Voting Panel consisted of 27 people, including rheumatologists, community rheumatologists, pulmonologists, community pulmonologists, and 3 patient representatives. The role of the Voting Panel was to vote on the drafted recommendation statements derived from the PICO questions, keeping the evidence report, their expertise and experience, and patient values and preferences in mind.

The ACR provided training for everyone involved in the development of this guideline, which included explanations of the ACR guideline process and GRADE methodology. See Supplementary Materials 4 for team/panel rosters.

Patient Panel

The Patient Panel, consisting of 5 adult men and 16 adult women who are at risk for or have been diagnosed with ILD, was convened on February 8, 2023. Dr. Marcy Bolster, Dr. Michael George, and Dr. Reza Mirza, members of the Core Team, and one ACR staff person facilitated the four-hour webinar discussion. The participants were first presented with the background and scope of the guideline project. The Patient Panel reviewed the evidence synthesized by the Literature Review Team as several PICO questions were discussed. The participants were encouraged to consider their personal experiences relevant to the questions and judge the importance of the outcomes and vote on the drafted recommendation statements accordingly. Three patients on the Voting Panel, who had been at the Patient Panel meeting, presented the values and preferences of the Patient Panel and their voting results to the Voting Panel during two four-day Voting Panel meetings held February 28-March 1, 2023 and March 29-30, 2023.

Disclosures and Management of Conflicts of Interest

Per ACR policy, everyone who was intellectually involved in the project (i.e., considered for guideline authorship) was required to disclose all relationships (https://www.rheumatology.org/Practice-Quality/Clinical-Support/Clinical-Practice-Guidelines/Integrative-RA-Treatment). Disclosures were evaluated to determine if any relationships were considered potential conflicts of interest for purposes of this project. Individuals whose primary employment (≥ 51% of work time/effort) was with a company that manufactured or sold therapeutics or diagnostics were not eligible to participate.

The project's principal investigator (PI) and the Literature Review Team leader had no relevant conflicts of interest for the full 12 months before this project began, and a majority of guideline development team members had no relevant conflicts of interest for the duration of the project.

Intellectual conflicts, such as a prior publication or scientific presentation on ILD, were recognized as important and were required to be disclosed, but because they were ubiquitous, intellectual conflicts were not counted as conflicted toward the allowed threshold.

Participant disclosures were initially shared in the project plan, which was posted online for public comment as the project began. Disclosures were updated and shared again with each project participant via email prior to the Voting Panel meeting. Updated participant disclosures are included online with this manuscript.

Scope and Target Audience

The scope of this project included the development of evidence-based recommendations for clinicians who care for people with systemic autoimmune rheumatic disease who are at risk for or have been diagnosed with interstitial lung disease (ILD).

The target audience for this guideline includes adults with systemic autoimmune rheumatic disease and their health care providers. Derivative products may be developed in the future to facilitate implementation of this guideline to these audiences.

Establishing Key Principles and PICO Development

The Core Leadership Team collaborated with Literature Review Team and Voting Panel members to develop the initial set of PICO-formatted clinical questions for the guideline, as well as identify pre-specified outcomes that were considered critical for each PICO question (see Supplementary Materials 2).

The Core Leadership Team held weekly conference calls, convened an initial virtual meeting of the Core Leadership Team, Literature Review Team, and Voting Panel in which the scope of the guideline was determined, and then developed the PICO questions. The PICO questions were posted for 30 days on the ACR website for public comment and revised accordingly.

Systematic Synthesis of the Literature

Population of Interest

The population of interest included:

- Individuals aged ≥17 years
- Rheumatoid arthritis (RA), Systemic sclerosis (Scleroderma, SSc), Mixed Connective Tissue Disease (MCTD), Polymyositis, Dermatomyositis, MDA5 Dermatomyositis, Immune-Mediated Necrotizing Myositis, Antisynthetase syndrome, Sjogren's syndrome (screening questions)

- Diagnosed with ILD or progression of ILD (treatment and monitoring questions)

Exclusion criteria were:

- Individuals aged ≤16 years
- Juvenile scleroderma, juvenile systemic sclerosis, juvenile dermatomyositis, juvenile idiopathic arthritis, Sarcoidosis, Interstitial Pneumonia with Autoimmune Features (IPAF), ankylosing spondylitis, ANCA-associated vasculitis, Systemic lupus erythematosus, Undifferentiated connective tissue disease
- Idiopathic Pulmonary Fibrosis
- Idiopathic interstitial pneumonia
- Unclassifiable ILD
- Overlap syndromes (e.g., systemic scleroderma [SSc] + myositis, RA + SSc, etc.)
- Other populations, not mentioned in the "Included" section.

Critical outcomes

The critical outcomes, as prespecified by the Core Team, include the following:

- **Critical outcomes for screening questions:**
    - Diagnostic accuracy
    - Disease-related outcomes*
    - Diagnostic testing-related adverse events
- **Critical outcomes for monitoring questions:**
    - Responsiveness/sensitivity to change in the test
    - Disease-related outcomes*
    - Treatment-related serious adverse events†
    - Testing-related adverse events†
- **Critical outcomes for medical management questions**:
    - Disease-related outcomes*
    - Treatment-related adverse events†

*Disease-related outcomes included mortality, disability, and health-related quality of life.
†Adverse events of interest included serious adverse events, toxicity leading to discontinuation, and other adverse reactions.

Surrogate outcomes were disease activity/disease progression defined by forced vital capacity (FVC), diffusion capacity for carbon monoxide (DLCO), CT thorax: the extent of disease, and disease progression.

Interventions

The following interventions were within the scope of this review:

- Pulmonary Function Tests (PFTs)
- History/physical alone (e.g., shortness of breath (dyspnea), functional class, and physician examination: crackles on auscultation)
- High-resolution CT Thorax
- 6-minute walk test distance
- Ambulatory desaturation
- Chest radiograph (chest x-ray)
- Bronchoscopy (may include broncho-alveolar lavage, transbronchial biopsy)
- Surgical lung biopsy
- csDMARDs: methotrexate, leflunomide, azathioprine, cyclophosphamide, mycophenolate, calcineurin inhibitors (tacrolimus, cyclosporine)
- bDMARDs: TNF inhibitors (etanercept, adalimumab, infliximab, golimumab, certolizumab pegol), IL-6 receptor antagonists (tocilizumab, sarilumab), anti-CD20 antibody (rituximab, ocrelizumab, obinutuzumab, ofatumumab), abatacept
- tsDMARDs: JAK inhibitors (tofacitinib, baricitinib, upadacitinib)
- Others: Oral prednisone, intravenous methylprednisolone, intravenous immunoglobulin (IVIG), plasma exchange (plasmapheresis)
- Antifibrotics: Pirfenidone, Nintedanib
- Stem cell transplant (autologous, mesenchymal stem cells, hematopoietic, myeloablative, non-myeloablative)
- Lung Transplant

The following interventions were outside of the scope of this review:

- Vaccines: influenza; COVID-19; Measles, Mumps, and Rubella (MMR); pneumococcus vaccine
- Education (self-management of oxygen, ILD disease)
- Physiotherapy (chest physiotherapy, airway clearance, incentive spirometry), Exercise (aerobic, resistance training, yoga, tai chi), Pulmonary Rehabilitation (cardio-pulmonary rehabilitation, resistance training, in a center versus home)
- Oxygen (oxygen desaturation at rest, oxygen desaturation <88% with exercise)
- Palliative care (cough, pain, air hunger, end-stage, end-of-life planning, when to initiate, what to initiate)
- Smoking cessation
- Fundoplication
- Gastrointestinal Medications: proton pump inhibitors, H2 blockers, promotility agents
- Ibritumomab (is anti-CD20, but it is radioimmunotherapy)
- Basiliximab
- Other interventions not mentioned in the "Included" section above.

Study Design

For all questions related to monitoring and management, we used a best evidence approach in which randomized (RCTs) or non-randomized controlled trials were considered as first line evidence. In the absence of controlled trials, we considered evidence from other study designs (e.g., observational studies) that reported on the population and intervention of interest. To capture adverse events, we also considered open-label extension studies of RCTs or other longitudinal observational studies that focused on safety and tolerability. For questions that focus on assessing screening accuracy, we included studies without a control group, specifically cohort and cross-sectional studies. We also included existing systematic reviews and guidelines from other societies only to confirm that we have included all relevant references.

Information Sources, and Search Strategy

A search was conducted and updated on August 1, 2022, using the following databases: Ovid MEDLINE(R) and Epub Ahead Of Print, In-Process, In-Data Review & Other Non-Indexed Citations, Daily And Versions(R) (searched from 1946 to May 3, 2021); Ovid MEDLINE(R) ALL (Original Search: 1946 to 08/01/2022); Ovid Embase (Original Search: 1974 to 08/01/2022). The search strategies can be found in **Supplementary Materials 5**. All searches were updated on January 6, 2023. We also supplemented database search with articles suggested by the Core Team members.

Study Selection

Two investigators independently assessed titles and abstracts of articles for potential inclusion. Subsequently, the full texts of these articles were obtained and independently evaluated by two investigators. To determine the final list of included or excluded articles, two more investigators reviewed the remaining full-text articles. In case of any disagreement, a consensus was reached through discussion (see Supplementay Materials 6).

Quality Assessment and Data Abstraction

We created a standardized table for data extraction in which one reviewer was responsible for extracting and evaluating data such as author, publication year, country, study type, patient characteristics, intervention type, and outcome data. For RCTs, data were obtained for control and intervention groups.

One researcher assessed individual study risk of bias (ROB). The Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool[1] was utilized to evaluate the ROB of diagnostic accuracy studies, while the Cochrane tool was employed to assess the ROB for RCTs (via GRADEPro)[2,3] and other study designs. A second review team member verified the accuracy of the extracted data and the ROB assessment to ensure consistency and reliability.

Certainty of Evidence Assessment

The certainty of the evidence was evaluated for each outcome using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system,[4] which assigns one of four grades reflecting the level of confidence in the effect estimate: high, moderate, low, or very low. The initial quality assessment is based on the study design: RCTs start at high quality, while observational studies start at low quality. The quality of evidence was downgraded based on five factors: risk of bias, inconsistency,

indirectness, imprecision, and publication bias. We used the GRADE guidelines for rating the quality of evidence. The quality of the evidence was assessed by one reviewer. A second reviewer from the literature review team verified the accuracy.

Of a particular note is the rating of evidence for surrogate outcomes. In alignment with GRADE recommendations, the Core Team characterized surrogate outcomes as those potentially linked to a clinically significant endpoint (e.g., mortality, functional improvement), but without necessarily correlating directly with it. For instance, forced vital capacity (FVC) and diffusing capacity for carbon monoxide (DLCO) served as surrogate outcomes in this review. When assessing surrogate outcomes, we reduced the certainty of patient-important outcomes (such as symptoms and mortality) by one or two levels to account for indirectness.

Presentation of Effects

Treatment effects for binary outcomes were calculated and presented as both relative and absolute effects with a random effects model when the meta-analysis was used. The effect for continuous outcomes was calculated using mean difference. All meta-analyses were conducted in RevMan.[5]

Relative effects convey the difference between the intervention and control groups in proportional terms. For instance, a 10% event rate in controls and a 5% event rate in the intervention equates to a 50% relative risk reduction ((10% - 5%) / 10%). Meanwhile, the same difference corresponds to a 5% absolute risk reduction (10% - 5% = 5%). Generally, absolute effects hold greater significance for patients.

In the tables, relative effects for dichotomous outcomes are expressed as either relative risk (RR) or odds ratio (OR). RR is the default effect size due to its ease of interpretation. However, in certain cases, RRs can result in implausible numbers when calculating absolute risk differences, prompting the use of ORs as an alternative to RRs.

Evidence Report Formulation

RevMan files were exported into GRADEpro software (https://www.gradepro.org/) to formulate a GRADE Summary of Findings (SoF) table for each PICO question (2), when possible. The quality of evidence for each outcome was evaluated by one literature review team member, then verified by the literature review leader (JT) using GRADE quality assessment criteria (1) with discordance resolved by discussion. The resulting SoF tables were compiled in an evidence report (see Supplementary Materials 3). The Core Leadership Team reviewed the evidence report prior to presentation to the Voting Panel.

Moving from Evidence to Recommendations

GRADE methodology specifies that voting panels make recommendations based on a consideration of the balance of benefits and harms/burdens of the treatment options under consideration, the quality of the evidence (i.e., confidence in the effect estimates), and patients' values and preferences. Key to the recommendation is the trade-off between desirable and undesirable outcomes; recommendations require estimating the relative value patients place on the outcomes.

A recommendation could be either in favor of or against the proposed intervention and either strong or conditional. According to GRADE, a recommendation is categorized as strong if the panel is very confident that the benefits of an intervention clearly outweigh the harms (or vice versa); a conditional recommendation denotes uncertainty regarding the balance of benefits and harms, such as when the evidence quality is low or very low, or when the decision is sensitive to individual patient preferences, or when costs are expected to impact the decision. Thus, conditional recommendations refer to decisions in which incorporation of patient preferences is a particularly essential element of decision-making. Judgments made in this guideline were based on the experience of the clinician panel members in shared decision making with their patients, on the experience and perspectives of this guideline's Patient Panel members and, to a considerable extent, on the results of discussion with the Patient Panel.

Consensus Building

The Voting Panel received the evidence report for review before it met to discuss and decide on the final recommendations. Individual online voting took place first, to ascertain any existing consensus on drafted recommendation statements that were based on the PICOs. This process was followed by two 2-day virtual webinar meetings of the Voting Panel, where they reviewed the evidence, edited recommendation statement wording, and provided final votes on the direction and strength of each recommendation. The webinar voting process was conducted using Poll Everywhere software (www.polleverywhere.com). A 70% consensus was used as the threshold for a recommendation; if 70% consensus was not achieved during an initial vote, the panel members held additional discussions before re-voting until at least 70% consensus was achieved. Following the meetings, additional clarifying questions were discussed by email and related voting took place via online survey.

Final Review and Approval of the Manuscript by the ACR

In addition to journal peer reviews, the manuscript was reviewed by the ACR Guideline Subcommittee, the ACR Quality of Care Committee, and the ACR Board of Directors. These ACR oversight groups did not make or mandate that specific recommendations be made within the guideline, but rather, served as

peer reviewers.

Moving from Recommendations to Practice

These recommendations are designed to support health care providers who work with patients in selecting therapies. Health care providers and patients must take into consideration not only clinical phenotype and level of disease activity, but also comorbidities, response and tolerance of prior therapies, patient's values and preferences, and patient's functional status and functional goals in choosing the optimal therapy for an individual patient at the given point in treatment.

REFERENCES

1. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336(7650):924-6.
2. Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. J Clin Epidemiol 2011;64:395-400.
3. Andrews J, Guyatt G, Oxman AD, Alderson P, Dahm P, Falck-Ytter Y, et al. GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. J Clin Epidemiol. 2013;66(7):719-25.
4. Andrews JC, Schunemann HJ, Oxman AD, Pottie K, Meerpohl JJ, Coello PA, et al. GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. J Clin Epidemiol. 2013;66(7):726-35.