The Lancet Rheumatology

Composite of Relevant Endpoints for Sjögren's Syndrome (CRESS): development and validation of a novel outcome measure

--Manuscript Draft--

Manuscript Number:	TLRHEU-D-21-00059R2
Article Type:	Article (Original Research)
Keywords:	Primary Sjögren's syndrome; endpoint; randomised controlled trial; therapy
Corresponding Author:	Liseth de Wolff, M.D. University Medical Centre Groningen: Universitair Medisch Centrum Groningen Groningen, Groningen NETHERLANDS
First Author:	Suzanne Arends, PhD
Order of Authors:	Suzanne Arends, PhD
	Liseth de Wolff, MD
	Jolien F van Nimwegen, MD PhD
	Gwenny M. P. J. Verstappen, PhD
	Jelle Vehof, MD PhD
	Michele Bombardieri, MD PhD
	Simon J. Bowman, PhD FRCP
	Elena Pontarini, PhD
	Alan N. Baer, MD
	Marleen Nys, PhD
	Jacques-Eric Gottenberg, MD PhD
	Renaud Felten, PhD
	Neelanjana Ray, PhD
	Arjan Vissink, DMD MD PhD
	Frans G.M. Kroese, PhD
	Hendrika Bootsma, MD PhD
Manuscript Region of Origin:	NETHERLANDS
Abstract:	Background: Recent randomised controlled trials (RCTs) in primary Sjögren's syndrome (pSS) used ESSDAI as their primary endpoint. Given the heterogeneous and complex nature of pSS, it might be more appropriate to also assess other clinically relevant disease features. We therefore developed a novel composite endpoint for assessing treatment efficacy in pSS patients: the Composite of Relevant Endpoints for Sjögren's Syndrome (CRESS). Methods: A multidisciplinary expert team selected clinically relevant items and candidate measurements. Cut-off points for response to treatment were based on expert opinion, previously published data regarding minimal clinically important improvement (MCII) and trial data, primarily the single-centre ASAPIII trial. Next, CRESS was validated in three independent RCTs: two potentially positive trials with rituximab (TRACTISS) and abatacept (multinational trial), and one negative trial with tocilizumab (ETAP). Findings: Based on expert opinion, five complementary items were selected to assess response: systemic disease activity by ClinESSDAI<5, patient-reported symptoms by ESSPRI (decrease ≥1 point and/or ≥15%), tear gland item by Schirmer's test or ocular staining score (in patients with abnormal Schirmer/OSS at baseline an increase ≥5 mm or decrease ≥2 points, respectively, in patients with normal values no change to abnormal for both), salivary gland item by unstimulated whole saliva secretion (increase ≥25%) or salivary gland ultrasonography (decrease ≥25%), and serological

	item by rheumatoid factor (decrease ≥25%) or immunoglobulin G (decrease ≥10%). Total CRESS response is defined as response on ≥3 of 5 items. At the primary endpoint visits, CRESS response rates were in ASAP-III: 24/40 (60%) for abatacept vs. 7/39 (18%) for placebo (p<0·0001), TRACTISS: 33/67 (49%) rituximab vs. 20/66 (30%) placebo (p=0·026), multinational abatacept: 41/92 (45%) abatacept vs. 30/95 (32%) placebo (p=0·067) and ETAP: 10/55 (18%) tocilizumab vs. 13/55 (24%) placebo (p=0·482). Interpretation: The CRESS is a feasible, well-balanced, composite endpoint for use in trials in pSS. Funding: None.

Composite of Relevant Endpoints for Sjögren's Syndrome (CRESS): development and validation of a novel outcome measure

Authors

Suzanne Arends*, Liseth de Wolff*, Jolien F. van Nimwegen, Gwenny M.P.J. Verstappen, Jelle Vehof, Michele Bombardieri, Simon J. Bowman, Elena Pontarini, Alan N. Baer, Marleen Nys, Jacques-Eric Gottenberg, Renaud Felten, Neelanjana Ray, Arjan Vissink, Frans G.M. Kroese, Hendrika Bootsma

**Both authors contributed equally*

University Medical Center Groningen, University of Groningen, Department of Rheumatology and Clinical Immunology, Groningen, The Netherlands

(S Arends PhD, L de Wolff MD, J F van Nimwegen PhD, G M P J Verstappen PhD, Prof. F G M Kroese PhD, Prof. H Bootsma PhD)

University Medical Center Groningen, University of Groningen, Department of Ophthalmology, Groningen, The Netherlands

(J Vehof PhD)

Queen Mary University of London, William Harvey Research Institute, Centre for Experimental Medicine and Rheumatology, London, United Kingdom

(Prof. M Bombardieri PhD, E Pontarini PhD)

Queen Elizabeth Hospital, Department of Rheumatology, Birmingham, United Kingdom

(Prof. S J Bowman PhD)

John Hopkins University School of Medicine, Department of Rheumatology, Department of Medicine, Baltimore, Maryland, USA

(A N Baer MD)

Bristol Myers Squibb, Belgium

(M Nys PhD)

CHU Strasbourg, Centre National de Référence des maladies auto-immunes et systémiques rare Est/Sud-Ouest (RESO), Department of Rheumatology, Strasbourg, Alsace, France

(Prof. J E Gottenberg PhD, R Felten MD)

Bristol Myers Squibb, Lawrenceville NJ, United States of America

(N Ray PhD)

University Medical Center Groningen, University of Groningen, Department of Oral and Maxillofacial Surgery , Groningen, The Netherlands

(Prof. A Vissink PhD)

Corresponding author
Prof. dr. H. Bootsma, MD, PhD, h.bootsma@umcg.nl
University Medical Center Groningen
Department of Rheumatology and Clinical Immunology
P.O. Box 30 001, 9700 RB Groningen, The Netherlands
Tel: +31503613432 / Fax: +31503619308

Word count: 3994

Figures: 3

Tables: 3

ABSTRACT

Background: Recent randomised controlled trials (RCTs) in primary Sjögren's syndrome (pSS) used ESSDAI as their primary endpoint. Given the heterogeneous and complex nature of pSS, it might be more appropriate to also assess other clinically relevant disease features. We therefore developed a novel composite endpoint for assessing treatment efficacy in pSS patients: the Composite of Relevant Endpoints for Sjögren's Syndrome (CRESS).

Methods: A multidisciplinary expert team selected clinically relevant items and candidate measurements. Cut-off points for response to treatment were based on expert opinion, previously published data regarding minimal clinically important improvement (MCII) and trial data, primarily the single-centre ASAPIII trial. Next, CRESS was validated in three independent RCTs: two potentially positive trials with rituximab (TRACTISS) and abatacept (multinational trial), and one negative trial with tocilizumab (ETAP).

Findings: Based on expert opinion, five complementary items were selected to assess response: systemic disease activity by ClinESSDAI<5, patient-reported symptoms by ESSPRI (decrease ≥1 point and/or ≥15%), tear gland item by Schirmer's test or ocular staining score (in patients with abnormal Schirmer/OSS at baseline an increase ≥5 mm or decrease ≥2 points, respectively, in patients with normal values no change to abnormal for both), salivary gland item by unstimulated whole saliva secretion (increase ≥25%) or salivary gland ultrasonography (decrease ≥25%), and serological item by rheumatoid factor (decrease ≥25%) or immunoglobulin G (decrease ≥10%). Total CRESS response is defined as response on ≥3 of 5 items.

At the primary endpoint visits, CRESS response rates were in ASAP-III: 24/40 (60%) for abatacept vs. 7/39 (18%) for placebo (p<0·0001), TRACTISS: 33/67 (49%) rituximab vs. 20/66 (30%) placebo (p=0·026), multinational abatacept: 41/92 (45%) abatacept vs. 30/95 (32%) placebo (p=0·067) and ETAP: 10/55 (18%) tocilizumab vs. 13/55 (24%) placebo (p=0·482).

Interpretation: The CRESS is a feasible, well-balanced, composite endpoint for use in trials in pSS.

Funding: None.

KEYWORDS

Primary Sjögren's syndrome, endpoint, randomised controlled trial, therapy

RESEARCH IN CONTEXT

Evidence before this study

Currently, no systemic immunomodulatory treatments have been registered for use in primary Sjögren's syndrome (pSS). Several, randomised controlled trials (RCTs) studying efficacy of various drugs in pSS failed to reach their primary endpoint. In most recent RCTs in pSS the European League Against Rheumatism (EULAR) Sjögren's Syndrome Disease Activity Index (ESSDAI) was used as the primary endpoint. We searched PubMed and EULAR/ACR congress abstract archives with the terms "Sjögren's syndrome", "randomised controlled trial" and "ESSDAI" up to January 2021 to identify RCTs of various immunomodulatory therapies using ESSDAI as primary endpoint, and how they performed. We identified eight RCTs: two phase II RCTs (prezalumab and seletalisib) failed, whereas three phase II RCTs (iscalimab, leflunomide/hydroxychloroquine and ianalumab) did find significant effects of active treatment compared to placebo. These results still need to be confirmed in phase III RCTs. Three phase III RCTs (two in abatacept, one in tocilizumab) failed, showing large response on ESSDAI (at group level more than the minimal clinically important improvement (MCII) of ≥3 points compared to baseline) in both active and placebo-treated patients. Given the heterogeneous nature of pSS, there is need for a composite endpoint including multiple clinically relevant aspects. Such a composite endpoint is presumed to be more appropriate in demonstrating drug efficacy than one that reports on a single aspect of this heterogeneous disease.

Added value of this study

A new composite endpoint, the Composite of Relevant Endpoints for Sjögren's Syndrome (CRESS) was developed to be used in future clinical trials. The CRESS consists of five complementary, clinically relevant items: a systemic disease activity, patient-reported symptoms, tear gland, salivary gland, and serological item. With the CRESS, higher response rates in abatacept and rituximab treated patients compared to placebo were found in RCTs which previously showed negative primary endpoint results. CRESS shows lower placebo response rates compared to the ESSDAI MCII, which is essential to be able to demonstrate treatment efficacy.

Implications of all the available evidence

The CRESS is suitable for use as primary endpoint in future clinical trials in pSS, particularly because this comprehensive tool includes a combination of disease activity, functional and serological parameters which are all important dimensions of this heterogeneous, systemic

auto-immune disease. This is crucial for the search of new, effective therapies for a broad selection of pSS patients.

INTRODUCTION

Primary Sjögren's syndrome (pSS) is a systemic autoimmune disease with a highly heterogeneous presentation. Main symptoms are ocular and oral sicca, fatigue and arthralgias. Many systemic symptoms can manifest, including arthritis, renal, cutaneous and pulmonary involvement and peripheral neuropathy. Additionally, laboratory abnormalities such as lymphocytopenia, hypergammaglobulinemia and presence of autoantibodies, including anti-SSA/SSB and rheumatoid factor (RF), can occur.[1]

Treatment currently relies on symptomatic treatment of sicca complaints and broad-spectrum immunosuppression, since no immunomodulatory systemic therapies have been registered for use in pSS patients.[2] Effective therapy is highly needed and many promising biological drugs are in development.

Positive open-label studies with multiple immunomodulatory drugs have been conducted[3,4], but larger randomised controlled trials (RCTs) failed to prove clinical efficacy.[5–9] So far, a wide variety of single endpoints have been used, but were not able to show treatment efficacy. The most frequently used endpoint is currently the European League Against Rheumatism (EULAR) Sjögren's Syndrome Disease Activity Index (ESSDAI). Recent RCTs using ESSDAI as primary endpoint showed large response rates both in the active treatment and placebo group[5,6], which could explain why these trials did not meet their efficacy endpoint.

A primary study endpoint that is feasible, valid, reliable, clinically relevant, sensitive to change, and able to discriminate between active treatment and placebo is crucial to prove clinical efficacy. Recent results give rise to the question whether the ESSDAI is fit to be used as primary endpoint on its own.[10] ESSDAI only measures systemic disease activity while there are other prominent and relevant pSS features, such as sicca symptoms, fatigue and decreased saliva and tear production. Moreover, although the ESSDAI includes a biological domain, additional serological markers are used in daily clinical practice to monitor disease activity. The hypothesis is that, given the complexity of pSS, a composite endpoint including multiple clinically relevant aspects might be more appropriate than a single endpoint. This is in line with the use of composite measures in other heterogeneous, immune-mediated diseases.[11,12] Furthermore, the use of a composite endpoint may facilitate inclusion of a broad selection of patients with different levels of disease activity and manifestations. A combined response at multiple items may also lower placebo response rates, facilitating discrimination.

Therefore, the objective of this study was to develop and validate a composite endpoint for assessing treatment efficacy in pSS based on expert opinion and trial data: the Composite of Relevant Endpoints for Sjögren's Syndrome (CRESS).

METHODS

The following steps were taken for development of the CRESS. 1) A multidisciplinary team was set up, consisting of experts in the pSS field, including a rheumatologist (HB), ophthalmologist (JV), oral medicine expert (AV), immunologist (FK), pharmacologist (GV), clinical trialist (NR) and epidemiologist (SA). This team selected clinically relevant items to include in a composite endpoint. 2) For each item, experts selected candidate measurements based on clinical relevance and practical feasibility. 3) Definition of response for CRESS items was based on clinical relevance according to the experts, previously published data regarding minimal clinically important improvement (MCII), and trial data, primarily the Abatacept Sjögren Active Patients phase III (ASAP-III) trial[5], and for some items rituximab trials.[4,13] 4) CRESS response was assessed in the ASAP-III trial, including the balance between CRESS items and criterion validity with physician global disease activity (PhyGDA) as comparator. Finally, CRESS was validated in three independent RCTs.[6–8]

Development trial data

CRESS was developed using data from three trials conducted in the multidisciplinary tertiary referral expertise centre for pSS at the University Medical Center Groningen (UMCG, Groningen, Netherlands). ASAP-III is a single-centre, phase III, randomised, double-blind, placebo-controlled, trial in 80 pSS patients, treated with weekly subcutaneous injections of abatacept (125 mg) or placebo for 24 weeks.[5] Data from the intention to treat population at 24 weeks were used (abatacept n=40, placebo n=39). Furthermore, two rituximab trials, an RCT[13] and an open-label trial[4], were combined to 31 unique pSS patients treated with rituximab (1000 mg intravenous infusion at day 1 and 15). For baseline characteristics of ASAP-III and rituximab trials see appendix p 3.

Statistical analyses of CRESS development

All statistical analyses were performed using IBM SPSS Statistics version 23.0. Receiver operating characteristic (ROC) analysis was used to assess discrimination between treatment groups in ASAP-III (abatacept versus placebo) for absolute and/or relative improvement in the selected candidate measurements and reported as area under the curve (AUC) with 95% confidence interval (CI). When possible, cut-off points were derived from the ROC analysis, based on the highest sum of sensitivity and specificity. Furthermore, cut-off points were based on expert opinion, taking measurement variation into account. Definition of total CRESS response was selected by calculating number of responders on one to five of the individual items and based on optimal discrimination between abatacept and placebo treatment. To explore balance between CRESS items, the number of CRESS responders

who were responders on single items was calculated, and the number of responders was calculated when leaving out single items one by one.

External validation

CRESS was externally validated in three multi-centre, phase III, randomised, double-blind, placebo-controlled trials: rituximab (Trial of Anti-B cell Therapy in Patients with Primary Sjögren's Syndrome, TRACTISS) (n=133)[8], multinational abatacept (n=187)[6] and tocilizumab (Efficacy of TocilizumAb in Primary Sjögren's Syndrome, ETAP) (n=110)[7] using available data on CRESS items. The TRACTISS and multinational abatacept trial were identified as potentially positive trials, and the ETAP trial as negative trial since clinical, patient-reported, glandular and immunological outcome measures showed negative results.[7]

Baseline characteristics and treatment regimens can be found in the original publications.[6–8] In all trials, number and percentage of responders were calculated for separate CRESS items and total CRESS at their primary endpoint visit (week 48 for TRACTISS, week 24 for the other trials). Additionally, early CRESS response was assessed at week 24 in the TRACTISS trial and week 12 for the other trials. Patients with <3 items available for evaluating CRESS response were imputed as non-responder. Significance of total CRESS response was tested using Chi-Square test. Total CRESS response was compared to the previously validated MCII of ≥3 points in ESSDAI.

RESULTS

Based on expert opinion, five complementary items were selected to be included in the CRESS: systemic disease activity, patient-reported symptoms, tear gland, salivary gland, and serology. PhyGDA was not selected as item since the use of specific tests for separate items was preferred to a global disease activity measurement. Candidate outcome measurements, selected by the multidisciplinary experts based on clinical relevance and feasibility, are presented in appendix p 4, including previously validated definitions of normal or low scores and validated cut-off points for improvement, when available.
In Table 1, selected items and definition of response per CRESS item are shown. An extensive glossary with CRESS instructions is shown in appendix p 5-8. Response rates of ROC and expert cut-off points for all items in the ASAP-III and rituximab trials (when available) are shown in appendix p 9-10.

Since ESSDAI is currently utilized as primary outcome measure in most recent RCTs, ESSDAI and Clinical ESSDAI (ClinESSDAI) were evaluated for the systemic disease activity item. ROC analyses of absolute and relative change showed no discrimination between abatacept and placebo treatment both for ESSDAI and ClinESSDAI (appendix p 11). Also, the MCII of ≥3 points decrease[14,15] showed no discrimination between treatment groups, with high response rates in the placebo group (>50%).
Next, patients were analysed for reaching the validated cut-off for low disease activity (<5 points)[14,15] at week 24. Response rates for ESSDAI<5 were 14/40 (35%) vs. 8/37 (22%) for abatacept and placebo treated patients, for ClinESSDAI<5 this was 18/40 (45%) vs. 10/37 (27%) (Table 2). ClinESSDAI low disease activity was preferred by the experts, because ClinESSDAI leaves out the biological domain, and the serological item is assessed separately in the CRESS. Of patients with high disease activity (ClinESSDAI≥14) at baseline, 8/22 (36%) abatacept patients versus 2/16 (13%) placebo patients reached ClinESSDAI low disease activity (<5) at week 24, indicating that reaching low disease activity for patients with high disease activity at baseline is possible and discriminating between treatment groups.

For the patient-reported symptoms item, EULAR Sjögren's Syndrome Patient Reported Index (ESSPRI) was selected for analysis by the experts. ROC analyses showed cut-off points similar to the previously validated MCII in ESSPRI: decrease of ≥1 point or ≥15% (appendix p 11).[14] Using this definition of response, response rates for abatacept and placebo treated patients were 23/40 (58%) vs. 8/36 (22%) (Table 2).

For the tear gland item, the experts selected Schirmer's test, measured without anesthesia, Ocular Staining Score (OSS), and tear breakup time (TBUT) for analyses, using mean scores of both eyes. Relative change for Schirmer's test as endpoint is unwanted, as small changes over time, i.e. <5 mm, most likely reflect normal variability. ROC analysis of Schirmer's test showed a cut-off point of +0·3 mm. Since this falls within measurement variation, the set definition for the minimal clinically important difference (MCID), ≥5 mm increase, was selected.[16] The OSS is a sum score ranging from 0 to 12, taking into account corneal and conjunctival staining scores, scored on a logarithmic scale, and extra points given for the presence of filaments, staining in the pupillary area, and patches of confluent staining.[17] Therefore, relative change in OSS as endpoint is mathematically not justifiable. ROC analysis of OSS was not discriminating between treatment groups and no cut-off point could be calculated. For OSS, the MCID has not been studied yet. Based on expert opinion and a clinical trial using OSS as endpoint, a cut-off of ≥2 points decrease was selected.[18] ROC analyses of TBUT were not discriminating between treatment groups (appendix p 11). Schirmer's test and OSS, but not TBUT, were selected, because they are complementary tests, measuring glandular function and ocular surface disease and are both part of the ACR-EULAR classification criteria.[19]

Since improvement within the normal range was considered not clinically relevant for Schirmer's test and OSS, separate definitions of response were developed for patients with normal and abnormal values at baseline. For patients with abnormal values of Schirmer's test (≤5 mm) at baseline, response was defined as ≥5 mm increase, or for patients with abnormal OSS values (≥3 points) at baseline, response was defined as ≥2 points decrease in OSS. For patients with normal scores at baseline, response was defined as no change to abnormal scores in both Schirmer's test and OSS. For response rates of all separate measurements see Table 2. Combining response definitions for Schirmer's test and OSS, response rates were 18/40 (45%) vs. 12/37 (32%) for abatacept and placebo treated patients.

For the salivary gland item, unstimulated whole saliva secretion (UWS), stimulated whole saliva secretion (SWS), and salivary gland ultrasonography (SGUS) were selected for analysis. For UWS and SWS relative change was explored because of wide inter-patient variation in secretion rates. ROC analysis for UWS showed a cut-off point of 0·0% (appendix p 11), which falls within measurement variation. Therefore, a stricter cut-off point of ≥25% increase was selected.[20] ROC analysis for SWS showed a cut-off point of +44·2%. In line with UWS, a cut-off point of ≥25% increase was also analysed. For patients with no saliva production (0 ml/min) at baseline, any increase in saliva secretion was defined as response.

Since UWS showed higher response rates in rituximab trials (appendix p 9), and is more commonly applied in pSS patients and more uniform and feasible to perform than SWS, UWS was included in the CRESS. ROC analyses for absolute and relative change in total Hocevar score (SGUS) showed somewhat better discrimination for relative change, with a cut-off point of -14·7% (appendix p 11).[21] To minimize response due to natural variation, the cut-off point was set at ≥25% decrease in total Hocevar score. For additional SGUS analyses see appendix p 13-14. Since UWS and SGUS are complementary tests, measuring gland function and structural changes in the glands, both were selected. When combining these measurements, response rates were 23/40 (58%) vs. 15/37 (41%) for abatacept and placebo treated patients (Table 2).

For the serological item, levels of RF, Immunoglobulin G (IgG), complement (C3 and C4), and lymphocyte count were considered as possible measurements by the experts. Relative change in serological markers was explored because of wide inter-patient variation in these scores. ROC analysis for RF showed a cut-off point of -23·0%, with good discrimination between treatment groups. Therefore, the cut-off point was set at ≥25% decrease. ROC analysis for IgG showed a cut-off point of -2·2% (appendix p 11). Because of the relatively small effect on IgG of abatacept (Table 2), and different biological drugs might induce different serological responses, IgG response was also evaluated in patients on rituximab (appendix p 9). The cut-off point of ≥10% was selected, giving acceptable sensitivity and specificity. Serum complement and lymphocyte count had no additional value (appendix p 15). Combining the complementary serological markers, RF and/or IgG response rates were 25/40 (63%) vs. 7/37 (19%) for abatacept and placebo treated patients (Table 2).

For a visual overview of CRESS items see Figure 1A. Response on ≥3 out of 5 items was most discriminating between treatment groups and therefore selected as definition of total CRESS response (Table 3). Total CRESS response rates were 24/40 (60%) vs. 7/39 (18%) for abatacept and placebo treated patients ($p<0·0001$) (Table 2, Figure 1B).
If OSS and/or SGUS are not available, the concise CRESS (cCRESS) can be used, leaving Schirmer's test and UWS for assessment of tear and salivary gland items, respectively. Total cCRESS response rates were 25/40 (63%) vs. 3/39 (8%) for abatacept and placebo treated patients ($p<0·0001$) (appendix p 16).
The number of CRESS responders who are responder on single items was well-balanced for all items (Figure 1C and appendix p 17). Criterion validity was confirmed by exploring the agreement of total CRESS response with PhyGDA (appendix p 18).

For external validation, the cCRESS was analysed in TRACTISS and ETAP trials, since OSS and SGUS Hocevar score were not available. The serological item in the ETAP trial was based on IgG only, since numeric RF values were not available. Furthermore, IgG was only measured at week 12 and imputed at week 24. In the multinational abatacept trial, the CRESS without SGUS was analysed.

Total CRESS or cCRESS response rates at the different time points are shown in Figure 2. Validation of the (c)CRESS in the TRACTISS and multinational abatacept trial as potentially positive trials demonstrated significant discrimination at the primary endpoint visit of the TRACTISS trial: 33/67 (49%) rituximab vs. 20/66 (30%) for placebo (p=0·026), and a trend towards significance for the multinational abatacept trial: 41/92 (45%) for abatacept vs. 30/95 (32%) for placebo (p=0·067). The ETAP trial was identified as a negative trial, which cCRESS confirmed with low response rates in both treatment groups: 10/55 (18%) for tocilizumab vs. 13/55 (24%) for placebo (p=0·482).

As shown in Figure 3, CRESS resulted in low response rates of placebo treated patients. In the multinational abatacept and ETAP trials, placebo response rates were 24-32% when using CRESS, compared to 58-64% when using ESSDAI MCII (Δ≥3). In the TRACTISS trial, CRESS was able to demonstrate a higher response rate for rituximab compared to placebo, whereas no distinction could be made based on ESSDAI MCII response between both treatment groups at primary endpoint visit. See appendix p 20-25 for external validation results of separate CRESS items and appendix p 26-27 for supportive analyses of selected CRESS items.

DISCUSSION

The choice of a primary study endpoint is a crucial step in designing clinical trials. Due to the heterogeneity and complexity of pSS, a composite endpoint is needed to reflect all important aspects of pSS. In the present study, we developed and validated the CRESS, a feasible composite endpoint combining disease activity, functional, and serological parameters, which enables discrimination between active treatment and placebo.

Our multidisciplinary expert team selected five complementary, clinically relevant items for the CRESS: systemic disease activity, patient-reported symptoms, tear gland, salivary gland, and serology. We found that all five CRESS items contributed equally to total CRESS response in the ASAP-III trial. When evaluating disease activity, systemic and patient-reported symptoms are both important. The validated instrument ESSDAI is widely used by physicians and researchers to assess systemic disease activity, and ESSPRI by patients to report prominent symptoms. High systemic disease activity may lead to severe complications of pSS, while patient-reported symptoms have greater impact on quality of life (QoL) of pSS patients.[22,23] Previous studies have demonstrated that correlation between ESSDAI and ESSPRI is poor.[24] This indicates that these indices are complementary to each other, which is the reason to include both in the CRESS. ClinESSDAI low disease activity was selected for the systemic disease activity item, since this was more discriminating and might be more relevant for patients than the ClinESSDAI MCII ($\Delta \geq 3$). Low disease activity as response criterion has also been used in systemic lupus erythematosus. The Lupus Low Disease Activity State (LLDAS) has been developed and validated in several RCTs[25] and was found to be associated with reduced damage accrual and higher health-related QoL.[26,27] The ESSPRI, consisting of three questions regarding dryness, fatigue and pain, was preferred to more elaborate questionnaires assessing separate patient-reported features, which might be more appropriate to use as secondary endpoints in trials.

For the tear and salivary gland items, two complementary tests assessing glandular function and structural abnormalities were included: Schirmer's test and OSS for the tear gland and UWS and SGUS for the salivary gland. For Schirmer's test, variability is high, especially within normal scores.[28] Therefore, improvement in Schirmer's test was considered clinically relevant only for patients with abnormal values, thus definition of response was split for normal and abnormal values at baseline. UWS was preferred to SWS because UWS is also included as an ACR-EULAR classification criterion[19] and is more uniform and feasible to perform. In contrast, SWS is performed using gustatory (e.g. citric acid) or mechanical stimulation (chewing). SGUS was included because it is an up-coming, non-invasive tool for imaging, revealing structural changes of salivary glands. Because it is not yet clear which items of SGUS scores are most important in evaluation of therapeutic response and what

each of these items reflect, we selected the widely applied total Hocevar score for the CRESS. Previously, we found that reliability varies between individual items of the Hocevar score. Total Hocevar score showed excellent intra-observer and good inter-observer reliability, although total scores varied more between observers in patients with higher ultrasonographic scores.[29] Therefore, relative change in total Hocevar score was considered as most applicable, and the cut-off selected was ≥25%. Further research and international collaboration is needed to determine the optimal SGUS scoring system for assessment of therapeutic response.

We decided to include the ClinESSDAI and serology as separate items because the biological domain of the ESSDAI is too global, based on cut-offs for IgG, meaning that improvement in IgG can be undetected. Moreover, the biological domain does not include RF autoantibodies, whereas RF is an important marker for disease activity and has been associated with development of MALT.[30] RF may be evaluated as total RF or IgM-RF, depending on the assay method. The vast majority of total RF consists of IgM-RF and response on this serological item is defined using relative change. Therefore, the cut-off point of ≥25% can be applied among different assays as was also the case in the validation trials. Immunologic and kinetic differences between formation of IgM-RF and IgG provide a rationale for combined analysis in clinical trials; IgM-RF (shorter half-life) is mainly produced by short-lived plasma cells[31] and IgG antibodies (longer half-life) are mainly produced by long-lived plasma cells. This is illustrated by the modest effect of B-cell depletion therapy on IgG levels and persistence of Ig-producing cells in salivary gland tissue.[9,32]

In the two external validation RCTs that were identified as potentially positive trials, significant discrimination with cCRESS was seen at the primary endpoint visit in the TRACTISS trial (week 48: 49% vs. 30%), and a trend towards significance with CRESS was seen in the multinational abatacept trial (week 24: 45% vs. 32%). The ETAP trial was identified as negative trial, which cCRESS confirmed, with low response rates in both treatment groups (week 24: 18% vs. 24%).

In the external validation trials with high baseline ESSDAI as inclusion criterion (multinational abatacept and ETAP), CRESS was able to approximately half placebo response rates compared to ESSDAI MCII (24-32% vs. 58-64%). Patients without enough items available to evaluate CRESS response were imputed as non-responder. Even when excluding patients with <3 CRESS items available, discrimination between treatment groups remained similar, and placebo response in all trials remained low (27-35%). This is an important improvement, since limited placebo response is essential to demonstrate treatment efficacy.

In the TRACTISS trial, no early treatment effect was seen at week 24. A longer follow-up period of 48 weeks with repeated cycles of rituximab might be needed to establish a treatment effect. The relatively high placebo response at week 24 may be explained by the

high dose of glucocorticoids. When analysing the individual CRESS items, there was a high response for ClinESSDAI <5 in both treatment groups (week 48: 76% vs. 60%). This can be explained because TRACTISS did not apply moderate to high systemic disease activity at baseline as inclusion criterion, as opposed to the other trials. However, the alternative of assessing the (Clin)ESSDAI MCII leads to low response rates, since for patients with already low ESSDAI scores it would be impossible to improve, which restricts response to ≥3/4 items. As seen from our additional analyses, replacing ClinESSDAI low disease activity by ClinESSDAI MCII in the cCRESS led to less discrimination between treatment groups in the TRACTISS trial. From a patients' point of view, it is also important that drugs become available for patients with a high symptom burden and a relatively low systemic disease activity, which is often the case in newly diagnosed pSS patients. All recent RCTs which used the ESSDAI as primary endpoint, included patients with moderate to high systemic disease activity (ESSDAI≥5). With the CRESS, patients with a relatively low systemic disease activity can also be included in clinical trials.

Limitations to this study are that data on sensitivity to change of the CRESS in pSS patients with long-standing disease is not yet available, since RCTs included patients with a relatively short disease duration. If patients with long-standing disease and damaged glands are included, certain CRESS items might be less sensitive to change, especially the tear and salivary gland items, and the dryness component of the ESSPRI. However, objectively measured glandular function and sicca symptoms are not always coupled. Furthermore, it is unclear whether dryness can improve in damaged glands. Since symptoms due to damage are not scored in the ClinESSDAI, this item and the serological item are expected to remain sensitive to change.

The CRESS was developed foremost based on expert opinion. Another composite endpoint, the Sjögren's Syndrome Responder Index (SSRI), includes five domains (visual analogue score for fatigue, ocular and oral dryness, UWS, and ESR) and was developed in 2015 based on a more data-driven approach using data of rituximab trials.[33] This endpoint lacks important measures for pSS such as systemic disease activity and objective measures for tear gland function. Moreover, the SSRI response was inferior to CRESS response in the ASAP-III trial, with response rates at week 24 of 13/40 (33%) in the abatacept and 9/39 (23%) in the placebo group.[5] Currently, an Innovative Medicines Initiative (IMI) project, NECESSITY[34], is working on finding a new composite endpoint based on a data-driven approach of candidate criteria and a Delphi procedure, in which the CRESS will also be included for further analyses.

In conclusion, the CRESS is a feasible composite endpoint which consists of well-balanced, complementary, and clinically relevant items for all pSS patients. We showed that the CRESS is able to discriminate between active treatment and placebo in pSS patients and

thus demonstrate clinical efficacy. As next step, validation of the CRESS in a future, prospective RCT is warranted. This study is an important advance in the search for valid composite endpoints for use in future clinical trials in pSS.

STATEMENTS

Competing interests statement

S. Arends: no competing interests

L. de Wolff: no competing interests

J.F. van Nimwegen: speaker and consultant for Bristol Myers Squibb

G.M. Verstappen: no competing interests

J. Vehof: speaker and consultant for Santen and Tramedico

M. Bombardieri: unrestricted grants from Amgen/Medimmune and Janssen, consultant for Amgen/Medimmune, Janssen, GSK, UCB.

S.J. Bowman: receives part funding of his salary from the NIHR Birmingham Biomedical Research Centre, consultant for Novartis, Abbvie and Galapagos and Abbvie

E. Pontarini: no competing interests

A.N. Baer: consultant for Bristol-Myers Squibb and Sanofi-Aventis

J. Gottenberg: received honoraries and research grants from BMS and Pfizer and honoraries from CSL Behring, Lilly, Janssen, UCB, Roche

R. Felten: no competing interests

M. Nys: is an employee and shareholder of Bristol Myers Squibb

N. Ray: is an employee and shareholder of Bristol Myers Squibb

A. Vissink: unrestricted grants from Bristol Myers Squibb and Roche

F.G.M. Kroese: unrestricted grants from Bristol Myers Squibb, consultant for Bristol Myers Squibb, speaker for Bristol Myers Squibb, Roche and Janssen-Cilag

H. Bootsma: unrestricted grants from Bristol Myers Squibb and Roche, consultant for Bristol Myers Squibb, Roche, Novartis, Medimmune, Union Chimique Belge, speaker for Bristol Myers Squibb and Novartis

Contributors

SA, JV, GMV, AV, NR, FGMK, HB were part of the multidisciplinary expert team. SA and LW performed the statistical analysis for the development and validation, EP performed the statistical analysis for the validation in the TRACTISS trial, MN performed the statistical analysis for the validation in the multinational abatacept trial. SA and LW wrote the first draft of the article. All authors interpreted the data, contributed to writing, critically revised the article, and approved the final submitted version.

Acknowledgements

The authors would like to acknowledge valuable discussions with M. Nowak, D. Grasela and S. Banerjee of BMS and with K. Hammitt of the Sjögren's Foundation.

Trial contributors

The ASAP-III trial: E. Mossel, G. van Zuiden, K. Delli, A. Stel, B. van der Vegt, E. Haacke, L. Olie, L. Los, S. Pringle and F. Spijkervet; the TRACTISS trial: Leeds CTRU and the colleagues involved in the TRACTISS trial: C. Everett, J. O'Dwyer, P. Emery, C. Pitzalis, W. Ng, C. Pease, E. Price, N. Sutcliffe, N. Gendi, F. Hall, S. Ruddock, C. Fernandez, C. Reynolds, C. Hulme, K. Davies, C. Edwards, P. Lanyon, R. Moots, E. Roussou, I. Giles, L. Sharples; the multinational abatacept trial: E. St Clair, T. Sumida, T. Takeuchi, G. Foulks, S. Mukherjee, R. Wong, B. Gavin; and the ETAP trial: V. Devauchelle-Pensec, R. Seror, P. Duffau, D. Saadoun, E. Hachulla, PY. Hatron, C. Salliot, A. Perdriger, J. Morel, A. Mékinian, O. Vittecoq, JM. Berthelot, E. Dernis, V. Le Guern, P. Dieudé, C. Larroche, C. Richez, T. Martin, C. Zarnitsky, G. Blaison, P. Kieffer, F. Maurier, A. Dellal, S. Rist, E. Andres, A Contis, E. Chatelus, C. Sordet, J. Sibilia, C. Arnold, M. Tawk, O. Aberkane, L. Holterbach, P. Cacoub, A. Saraux, X. Mariette and N. Meyer.

Data sharing

Currently there are no plans to share additional data beyond what is shared in this article.

REFERENCES

1. Mariette X, Criswell LA. Primary Sjögren's syndrome. N Engl J Med. 2018;378(10):931–9.
2. Ramos-Casals M, Brito-Zerón P, Bombardieri S et al. EULAR recommendations for the management of Sjögren's syndrome with topical and systemic therapies. Ann Rheum Dis. 2020;79(1):3–18.
3. Meiners PM, Vissink A, Kroese FGM et al. Abatacept treatment reduces disease activity in early primary Sjögren's syndrome (open-label proof of concept ASAP study). Ann Rheum Dis. 2014;73(7):1393–6.
4. Meiners PM, Arends S, Brouwer E, Spijkervet FKL, Vissink A, Bootsma H. Responsiveness of disease activity indices ESSPRI and ESSDAI in patients with primary Sjögren's syndrome treated with rituximab. Ann Rheum Dis. 2012;71(8):1297–302.
5. van Nimwegen JF, Mossel E, van Zuiden GS et al. Abatacept treatment for patients with early active primary Sjögren's syndrome: a single-centre, randomised, double-blind, placebo-controlled, phase 3 trial (ASAP-III study). Lancet Rheumatol. 2020;9913(19):1–11.
6. Baer A, Gottenberg J-E, St.Clair EW et al. Efficacy and safety of abatacept in active primary Sjögren's syndrome: results of a phase III, randomised, placebo-controlled trial. Ann Rheum Dis. 2020;0:1–10.
7. Felten R, Devauchelle-Pensec V, Seror R et al. Interleukin 6 receptor inhibition in primary Sjögren syndrome: A multicentre double-blind randomised placebo-controlled trial. Ann Rheum Dis. 2020;0:1–10.
8. Bowman SJ, Everett CC, O'Dwyer JL et al. Randomized Controlled Trial of Rituximab and Cost-Effectiveness Analysis in Treating Fatigue and Oral Dryness in Primary Sjögren's Syndrome. Arthritis Rheumatol. 2017;69(7):1440–50.
9. Devauchelle-Pensec V, Mariette X, Jousse-Joulin S et al. Treatment of primary sjögren syndrome with rituximab a randomized trial. Ann Intern Med. 2014;29(1):6–12.
10. de Wolff L, Arends S, van Nimwegen JF, Bootsma H. Ten years of the ESSDAI: is it fit for purpose? Clin Exp Rheumatol. 2020;38(4):283–90.
11. Rider L.G., Aggarwal R., Machado P.M.,et al. Update on outcome assessment in myositis. Nat Rev Rheumatol. 2018 ;14(5) :303-18.
12. Khanna D, Huang S, Lin CJF, Spino C. New composite endpoint in early diffuse cutaneous systemic sclerosis: revisiting the provisional American College of Rheumatology Composite Response Index in Systemic Sclerosis. Ann Rheum Dis. 2020;0:1–10.

13. Meijer JM, Meiners PM, Vissink A et al. Effectiveness of rituximab treatment in primary sjögren's syndrome: A randomized, double-blind, placebo-controlled trial. Arthritis Rheum. 2010;62(4):960–8.
14. Seror R, Bootsma H, Saraux A et al. Defining disease activity states and clinically meaningful improvement in primary Sjögren's syndrome with EULAR primary Sjögren's syndrome disease activity (ESSDAI) and patient-reported indexes (ESSPRI). Ann Rheum Dis. 2016;75(2):382–9.
15. Seror R, Meiners P, Baron G et al. Development of the ClinESSDAI: A clinical score without biological domain. A tool for biological studies. Ann Rheum Dis. 2016;75(11):1945–50.
16. Wolffsohn JS, Arita R, Chalmers R et al. TFOS DEWS II Diagnostic Methodology report. Ocul Surf. 2017;15(3):539–74.
17. Whitcher JP, Shiboski CH, Shiboski SC et al. A Simplified Quantitative Method for Assessing Keratoconjunctivitis Sicca From the Sjögren's Syndrome International Registry. Am J Ophthalmol. 2010;149:405-415.
18. Boujnah Y, Mouchel R, El-Chehab H, Dot C, Burillon C, Kocaba V. Prospective, monocentric, uncontrolled study of efficacy, tolerance and adherence of cyclosporin 0.1 % for severe dry eye syndrome. J Fr Ophtalmol. 2018;41(2):129–35.
19. Shiboski CH, Shiboski SC, Seror R et al. 2016 American College of Rheumatology/European League Against Rheumatism classification criteria for primary Sjögren's syndrome: A consensus and data-driven methodology involving three international patient cohorts. Ann Rheum Dis. 2017;76(1):9–16.
20. Burlage FR, Pijpe J, Coppes RP et al. Variability of flow rate when collecting stimulated human parotid saliva. Eur J Oral Sci. 2005;113(5):386–90.
21. Hočevar A, Rainer S, Rozman B, Zor P, Tomšič M. Ultrasonographic changes of major salivary glands in primary Sjögren's syndrome. Evaluation of a novel scoring system. Eur J Radiol. 2007;63(3):379–83.
22. Lendrem D, Mitchell S, McMeekin P et al. Health-related utility values of patients with primary Sjögren's syndrome and its predictors. Ann Rheum Dis. 2014;73(7):1362–8.
23. Hammitt KM. Sjögren's: the patients' perspective. Clin Exp Rheumatol 2020;38(Suppl 126)25–6.
24. Seror R, Theander E, Brun JG et al. Validation of EULAR primary Sjögren's syndrome disease activity (ESSDAI) and Patient indexes (ESSPRI). Ann Rheum Dis. 2015;74(5):859–66.
25. Morand EF, Trasieva T, Berglind A, Illei GG, Tummala R. Lupus Low Disease Activity State (LLDAS) attainment discriminates responders in a systemic lupus erythematosus trial: Post-hoc analysis of the Phase IIb MUSE trial of anifrolumab. Ann

Rheum Dis. 2018;77(5):706–13.
26. Tsang-A-Sjoe MWP, Bultink IEM, Heslinga M, Voskuyl AE. Both prolonged remission and Lupus Low Disease Activity State are associated with reduced damage accrual in systemic lupus erythematosus. Rheumatology (Oxford). 2017;56(1):121–8.
27. Golder V, Kandane-Rathnayake R, Hoi AYB et al. Association of the lupus low disease activity state (LLDAS) with health-related quality of life in a multinational prospective study. Arthritis Res Ther. 2017;19(1):1–11.
28. Nichols KK, Mitchell GL, Zadnik K. The Repeatability of Clinical Measurements of Dry Eye. Cornea. 2004;23(3):272–85.
29. Delli K, Arends S, van Nimwegen JF et al. Ultrasound of the Major Salivary Glands is a Reliable Imaging Technique in Patients with Clinically Suspected Primary Sjögren's Syndrome. Ultraschall Med 2018;39:328-33.
30. Nocturne G, Virone A, Ng WF et al. Rheumatoid Factor and Disease Activity Are Independent Predictors of Lymphoma in Primary Sjögren's Syndrome. Arthritis Rheumatol. 2016;68(4):977–85.
31. William J, Euler C, Shlomchik MJ. Short-Lived Plasmablasts Dominate the Early Spontaneous Rheumatoid Factor Response: Differentiation Pathways, Hypermutating Cell Types, and Affinity Maturation Outside the Germinal Center. J Immunol. 2005;174(11):6879–87.
32. Hamza N, Bootsma H, Yuvaraj S et al. Persistence of immunoglobulin-producing cells in parotid salivary glands of patients with primary Sjögren's syndrome after B cell depletion therapy. Ann Rheum Dis. 2012;71(11):1881–7.
33. Cornec D, Devauchelle-Pensec V, Mariette X et al. Development of the Sjögren's Syndrome Responder Index, a data-driven composite endpoint for assessing treatment efficacy. Rheumatol (United Kingdom). 2015;54(9):1699–708.
34. New clinical endpoints in primary Sjögren's Syndrome: an interventional trial based on stratifying patients (NECESSITY). Innovative Medicines Initiative. Grant agreement number 806975.

Composite of Relevant Endpoints for Sjögren's Syndrome (CRESS): development and validation of a novel outcome measure~~a comprehensive tool for clinical trials~~

Authors

Suzanne Arends~~, PhD~~[1]*, Liseth de Wolff~~, MD~~[1]*, Jolien F. van Nimwegen~~, MD PhD~~[1], Gwenny M.P.J. Verstappen~~, PhD~~[1], Jelle Vehof~~, MD PhD~~[2], Michele Bombardieri~~, MD PhD~~[3], Simon J. Bowman~~, PhD FRCP~~[4], Elena Pontarini~~, PhD~~[3], Alan N. Baer~~, MD~~[5], Marleen Nys~~, PhD~~[6], Jacques-Eric Gottenberg~~, MD PhD~~[7], Renaud Felten~~, MD~~[7], Neelanjana Ray~~, PhD~~[8], Arjan Vissink~~, DMD MD PhD~~[9], Frans G.M. Kroese~~, PhD~~[1], Hendrika Bootsma~~, MD PhD~~[1]

Both authors contributed equally

[1]**University Medical Center Groningen, University of Groningen, Department of Rheumatology and Clinical Immunology, Groningen, The Netherlands**

(S Arends PhD, L de Wolff MD, J F van Nimwegen PhD, G M P J Verstappen PhD, Prof. F G M Kroese PhD, Prof. H Bootsma PhD)

[2]**University Medical Center Groningen, University of Groningen, Department of Ophthalmology, Groningen, The Netherlands**

(J Vehof PhD)

[3]**Queen Mary University of London, William Harvey Research Institute, Centre for Experimental Medicine and Rheumatology, London, United Kingdom**

(Prof. M Bombardieri PhD, E Pontarini PhD)

[4]**Queen Elizabeth Hospital, Department of Rheumatology, Birmingham, United Kingdom**

(Prof. S J Bowman PhD)

[5]**John Hopkins University School of Medicine, Department of Rheumatology, Department of Medicine, Baltimore, Maryland, USA**

(A N Baer MD)

[6]**Bristol Myers Squibb, Belgium**

(M Nys PhD)

[7]**CHU Strasbourg, Centre National de Référence des maladies auto-immunes et systémiques rare Est/Sud-Ouest (RESO), Department of Rheumatology, Strasbourg, Alsace, France**

(Prof. J E Gottenberg PhD, R Felten MD)

[8]**Bristol Myers Squibb, Lawrenceville NJ, United States of America**

(N Ray PhD)

[9]**University Medical Center Groningen, University of Groningen, Department of Oral and Maxillofacial Surgery , Groningen, The Netherlands**

(Prof. A Vissink PhD)

Corresponding author
Prof. dr. H. Bootsma, MD, PhD, h.bootsma@umcg.nl
University Medical Center Groningen
Department of Rheumatology and Clinical Immunology
P.O. Box 30 001, 9700 RB Groningen, The Netherlands
Tel: +31503613432 / Fax: +31503619308

Word count: 39944~~4731~~

Figures: 3

Tables: 3~~7~~

ABSTRACT

Background: Recent randomised controlled trials (RCTs) in primary Sjögren's syndrome (pSS) used ESSDAI as their primary endpoint. Given the heterogeneous and complex nature of pSS, it might be more appropriate to also assess other clinically relevant disease features. We therefore developed a novel composite endpoint for assessing treatment efficacy in pSS patients: the Composite of Relevant Endpoints for Sjögren's Syndrome (CRESS).

Methods: A multidisciplinary expert team selected clinically relevant items and candidate measurements. Cut-off points for response to treatment were based on expert opinion, previously published data regarding minimal clinically important improvement (MCII) and trial data, primarily the single-centre ASAPIII trial. Next, CRESS was validated in three independent RCTs: two potentially positive trials with rituximab (TRACTISS) and abatacept (multinational trial), and one negative trial with tocilizumab (ETAP).

Findings: Based on expert opinion, five complementary items were selected to assess response: systemic disease activity by ClinESSDAI<5, patient-reported symptoms by ESSPRI (decrease ≥1 point and/or ≥15%), tear gland item by Schirmer's test or ocular staining score (in patients with abnormal Schirmer/OSS at baseline an increase ≥5 mm or decrease ≥2 points, respectively, in patients with normal values no change to abnormal for both), salivary gland item by unstimulated whole saliva secretion (increase ≥25%) or salivary gland ultrasonography (decrease ≥25%), and serological item by rheumatoid factor (decrease ≥25%) or immunoglobulin G (decrease ≥10%). Total CRESS response is defined as response on ≥3 of 5 items.

At the primary endpoint visits, CRESS response rates were in ASAP-III: 24/40 (60%) for abatacept vs. 7/39 (18%) for placebo (p<0·0001), TRACTISS: 33/67 (49%) rituximab vs. 20/66 (30%) placebo (p=0·026), multinational abatacept: 41/92 (45%) abatacept vs. 30/95 (32%) placebo (p=0·067) and ETAP: 10/55 (18%) tocilizumab vs. 13/55 (24%) placebo (p=0·482).

Interpretation: The CRESS is a feasible, well-balanced, composite endpoint for use in trials in pSS.

Funding: None.~~See the original articles for funding of the included trials.~~

KEYWORDS

Primary Sjögren's syndrome, endpoint, randomised controlled trial, therapy

RESEARCH IN CONTEXT

Evidence before this study

Currently, no systemic immunomodulatory treatments have been registered for use in primary Sjögren's syndrome (pSS). Several, randomised controlled trials (RCTs) studying efficacy of various drugs in pSS failed to reach their primary endpoint. In most recent RCTs in pSS the European League Against Rheumatism (EULAR) Sjögren's Syndrome Disease Activity Index (ESSDAI) was used as the primary endpoint. We searched PubMed and EULAR/ACR congress abstract archives with the terms "Sjögren's syndrome", "randomised controlled trial" and "ESSDAI" up to January 2021 to identify RCTs of various immunomodulatory therapies using ESSDAI as primary endpoint, and how they performed. We identified eight RCTs: two phase II RCTs (prezalumab and seletalisib) failed, whereas three phase II RCTs (iscalimab, leflunomide/hydroxychloroquine and ianalumab) did find significant effects of active treatment compared to placebo. These results still need to be confirmed in phase III RCTs. Three phase III RCTs (two in abatacept, one in tocilizumab) failed, showing large response on ESSDAI (at group level more than the minimal clinically important improvement (MCII) of ≥3 points compared to baseline) in both active and placebo-treated patients. Given the heterogeneous nature of pSS, there is need for a composite endpoint including multiple clinically relevant aspects. Such a composite endpoint is presumed to be more appropriate in demonstrating drug efficacy than one that reports on a single aspect of this heterogeneous disease.

Added value of this study

A new composite endpoint, the Composite of Relevant Endpoints for Sjögren's Syndrome (CRESS) was developed to be used in future clinical trials. The CRESS consists of five complementary, clinically relevant items: a systemic disease activity, patient-reported symptoms, tear gland, salivary gland, and serological item. With the CRESS, higher response rates in abatacept and rituximab treated patients compared to placebo were found in RCTs which previously showed negative primary endpoint results. CRESS shows lower placebo response rates compared to the ESSDAI MCII, which is essential to be able to demonstrate treatment efficacy.

Implications of all the available evidence

The CRESS is suitable for use as primary endpoint in future clinical trials in pSS, particularly because this comprehensive tool includes a combination of disease activity, functional and serological parameters which are all important dimensions of this heterogeneous, systemic

auto-immune disease. This is crucial for the search of new, effective therapies for a broad selection of pSS patients.

INTRODUCTION

Primary Sjögren's syndrome (pSS) is a systemic autoimmune disease with a highly heterogeneous presentation. Main symptoms are ocular and oral sicca, fatigue and arthralgias. Many ~~other~~ systemic symptoms can manifest, including arthritis, renal, cutaneous and pulmonary involvement and peripheral neuropathy. Additionally, laboratory abnormalities such as lymphocytopenia, hypergammaglobulinemia and presence of autoantibodies, including anti-SSA/SSB and rheumatoid factor (RF), can occur.[1]

Treatment currently relies on symptomatic treatment of sicca complaints and broad-spectrum immunosuppression, since no immunomodulatory systemic therapies have been registered for use in pSS patients.[2] Effective therapy is highly needed and many promising biological drugs are in development.

Positive open-label studies with multiple immunomodulatory drugs have been conducted[3,4], but larger randomised controlled trials (RCTs) failed to prove clinical efficacy.[5–9] So far, a wide variety of single endpoints have been used, but ~~and none of them~~ were not able to show treatment efficacy. The most frequently used endpoint is currently the European League Against Rheumatism (EULAR) Sjögren's Syndrome Disease Activity Index (ESSDAI). ~~–~~ Recent RCTs using ~~the~~ ESSDAI as primary endpoint showed large response rates both in the active treatment and placebo group[5,6], which could explain why these trials did not meet their efficacy endpoint.

A primary study endpoint that is feasible, valid, reliable, clinically relevant, sensitive to change, and able to discriminate between active treatment and placebo is crucial to prove clinical efficacy. Recent results give rise to the question whether the ESSDAI is fit to be used as primary endpoint on its own.[10] ~~The~~ ESSDAI only measures systemic disease activity while there are other prominent and relevant pSS features, such as sicca symptoms, fatigue and decreased saliva and tear production. Moreover, although the ESSDAI includes a biological domain, additional serological markers are used in daily clinical practice to monitor disease activity. The hypothesis is that, given the complexity ~~nature~~ of pSS, a composite endpoint including multiple clinically relevant aspects might be more appropriate than a single endpoint. This is in line with the use of composite measures in other heterogeneous, immune-mediated diseases.[11,12] Furthermore, the use of a composite endpoint may facilitate inclusion of a broad ~~er~~ selection of patients with different levels of disease activity and manifestations. A combined response at multiple items may also lower placebo response rates, facilitating discrimination. ~~between active treatment and placebo, which is crucial to prove clinical efficacy.~~

Therefore, the objective of this study was to develop and validate a composite endpoint for assessing treatment efficacy in pSS based on expert opinion and trial data: the Composite of Relevant Endpoints for Sjögren's Syndrome (CRESS).

METHODS

The following steps were taken for development of the CRESS. 1) A multidisciplinary team was set up, consisting of experts in the pSS field, including a rheumatologist (HB), ophthalmologist (JV), oral medicine expert (AV), immunologist (FK), pharmacologist (GV), clinical trialist (NR) and epidemiologist (SA). This team selected clinically relevant items to include in a composite endpoint. 2) For each item, experts selected candidate measurements based on clinical relevance and practical feasibility. 3) Definition of response for CRESS items was based on clinical relevance according to the experts, previously published data regarding minimal clinically important improvement (MCII), and trial data, primarily the Abatacept Sjögren Active Patients phase III (ASAP-III) trial[5], and~~,~~ for some items~~,~~ rituximab trials.[4,13~~1~~] 4) CRESS response was assessed in the ASAP-III trial, including the balance between CRESS items and criterion validity with physician global disease activity (PhyGDA) as comparator. Finally, ~~the~~ CRESS was validated in three independent RCTs.[6–8]

Development trial data

CRESS was developed using data from three trials conducted in the multidisciplinary tertiary referral expertise centre for pSS at the University Medical Center Groningen (UMCG, Groningen, Netherlands). ASAP-III is a single-centre, phase III, randomised, double-blind, placebo-controlled, ~~phase 3~~ trial in 80 pSS patients, ~~, fulfilling the American-European Consensus Group (AECG) and American College of Rheumatology (ACR)-EULAR criteria,~~ treated with weekly subcutaneous injections of abatacept (125 mg) or placebo for 24 weeks.[5] Data from the intention to treat population at 24 weeks were used (abatacept n=40, placebo n=39). Furthermore, two rituximab trials, an RCT[13~~1~~] and an open-label trial[4], were combined to 31 unique pSS patients treated with rituximab (1000 mg intravenous infusion at day 1 and 15). For baseline characteristics of ASAP-III and rituximab trials see appendix p 3~~Supplementary Table 1~~.

Statistical analyses of CRESS development

All statistical analyses were performed using IBM SPSS Statistics version 23.0. ~~Analyses were performed using trial data at baseline and week 24.~~ Receiver operating characteristic (ROC) analysis was used to assess discrimination between treatment groups in ASAP-III (abatacept versus placebo) for absolute and/or relative improvement in the selected

candidate measurements and reported as area under the curve (AUC) with ~~a~~ 95% confidence interval (CI). When possible, cut-off points were derived from the ROC analysis, based on the highest sum of sensitivity and specificity. Furthermore, cut-off points were based on expert opinion, taking measurement variation into account. Definition of total CRESS response was selected by calculating number of responders on one to five of the individual items and based on optimal discrimination between abatacept and placebo treatment. To explore balance between CRESS items, the number of CRESS responders who were responders on single items was calculated, and~~. Also,~~ the number of responders was calculated when leaving out single items one by one.

External validation

~~The~~ CRESS was externally validated in three multi-centre, phase III, randomised, double-blind, placebo-controlled trials: rituximab (Trial of Anti-B cell Therapy in Patients with Primary Sjögren's Syndrome, TRACTISS) (n=133)[8], multinational abatacept (n=187)[6] and tocilizumab (Efficacy of TocilizumAb in Primary Sjögren's Syndrome, ETAP) (n=110)[7] using available data on CRESS items. The TRACTISS and multinational abatacept trial were identified as potentially positive trials, and the ETAP trial as ~~a~~ negative trial since clinical, patient-reported, glandular and immunological outcome measures showed negative results.[7]

Baseline characteristics and~~,~~ treatment regimens ~~and funding~~ can be found in the original publications.[6–8] In all trials, number and percentage of responders were calculated for separate CRESS items and total CRESS at their primary endpoint visit (week 48 for TRACTISS, week 24 for the other trials). Additionally, early CRESS response was assessed at week 24 in the TRACTISS trial and week 12 for the other trials. Patients with <3 items available for evaluating CRESS response were imputed as non-responder. Significance of total CRESS response was tested using ~~a~~ Chi-Square test. Total CRESS response was compared to the previously validated MCII of ≥3 points in ESSDAI.

RESULTS

~~Selection of items and candidate measurements~~

Based on expert opinion, five complementary items were selected to be included in the CRESS: systemic disease activity, patient-reported symptoms, tear gland, salivary gland, and serology. ~~Physician global disease activity (~~PhyGDA~~),~~ was not selected as item since the use of specific tests for ~~all contributing~~separate items was preferred to a global disease activity measurement. Candidate outcome measurements, selected by the multidisciplinary experts based on ~~their~~ clinical relevance and feasibility, are presented in appendix p 4~~Table 1~~, including previously validated definitions of normal or low scores and validated cut-off points for improvement, when available.

~~Definition of response and performance of items and measurements~~

In Table 1~~2~~, selected items and definition of response per CRESS item are shown. An extensive glossary with CRESS instructions is shown in ~~Supplementary Tables 2-5~~appendix p 5-8. Response rates of ROC and expert cut-off points for all items in the ASAP-III and rituximab trials (when available) are shown in ~~Supplementary Table 6~~appendix p 9-10.

~~*Systemic disease activity item*~~

Since ESSDAI is currently utilized as primary outcome measure in most recent RCTs, ESSDAI and Clinical ESSDAI (ClinESSDAI) were ~~selected~~evaluated ~~for evaluation~~ for the systemic disease activity item. ROC analyses of absolute and relative change showed no discrimination between abatacept and placebo treatment both for ESSDAI and ClinESSDAI (appendix p 11~~Table 3~~). Also, the MCII of ≥3 points decrease[14~~2~~,15~~3~~] showed no discrimination between treatment groups, with high response rates in the placebo group (>50%).

Next, patients were analysed for reaching the validated cut-off for low disease activity (<5 points)[14~~2~~,15~~3~~] at week 24. Response rates for ESSDAI<5 were 14/40 (35%) vs. 8/37 (22%) for abatacept and placebo treated patients, for ClinESSDAI<5 this was 18/40 (45%) vs. 10/37 (27%) (Table 2~~4~~). ClinESSDAI low disease activity was preferred by the experts, because ClinESSDAI leaves out the~~a~~ biological domain, and the serological item is assessed separately in the CRESS. ~~When focusing on~~Of patients with high disease activity (ClinESSDAI≥14) at baseline, 8~~/ of~~ 22 (36%) abatacept patients ~~treated with abatacept~~ versus 2~~/ of~~ 16 (13%) placebo patients ~~treated with placebo~~ reached ClinESSDAI low disease activity (<5) at week 24, indicating~~. This indicates~~ that reaching low disease activity for patients with ~~a~~ high disease activity at baseline is possible~~,~~ and ~~is~~ discriminating between ~~abatacept and placebo patients.~~treatment groups.

Patient-reported symptoms item

For the patient-reported symptoms item, EULAR Sjögren's Syndrome Patient Reported Index (ESSPRI) was selected for analysis by the experts. ROC analyses showed cut-off points similar to the previously validated MCII in ESSPRI: a decrease of ≥1 point or ≥15% (appendix p 11Table 3).[14,2] Using this definition of response, response rates for abatacept and placebo treated patients were 23/40 (58%) vs. 8/36 (22%) (Table 24).

Tear gland item

For the tear gland item, tThe experts selected Schirmer's test, measured without anesthesia, Ocular Staining Score (OSS), and tear breakup time (TBUT) for analyses, using mean scores of both eyes. Relative change for Schirmer's test as an endpoint is unwanted, as small changes over time, i.e. <5 mm, most likely reflect normal variability. ROC analysis of Schirmer's test showed a cut-off point of +0·3 mm. Since this falls within measurement variation, the set definition for the minimal clinically important difference (MCID), ≥5 mm increase, was selected.[16,4] The OSS is a sum score ranging from 0 to 12, taking into account corneal and conjunctival staining scores, scored on a logarithmic scale, and extra points given for the presence of filaments, staining in the pupillary area, and patches of confluent staining.[17,5] Therefore, relative change in OSS as an endpoint is mathematically not justifiable. ROC analysis of OSS was not discriminating between treatment groups and no cut-off point could be calculated. For OSS, the MCID has not been studied yet. Based on expert opinion and a clinical trial using OSS as endpoint, a cut-off of ≥2 points decrease was selected.[18] ROC analyses of TBUT were not discriminating between treatment groups (appendix p 11Table 3). Schirmer's test and OSS, but not TBUT, were selected, because they are complementary tests, measuring glandular function and ocular surface disease and are both part of the ACR-EULAR classification criteria.[19]

Since improvement within the normal range was considered not clinically relevant for Schirmer's test and OSS, separate definitions of response were developed for patients with normal and abnormal values at baseline. For patients with abnormal values of Schirmer's test (≤5 mm) at baseline, response was defined as ≥5 mm increase, or for patients with abnormal OSS values (≥3 points) at baseline, response was defined as ≥2 points decrease in OSS. For patients with normal scores at baseline, response was defined as no change to abnormal scores in both Schirmer's test and OSS. For rResponse rates offor all separate measurements seeare shown in Table 24. Combining response definitions for Schirmer's test

and OSS, response rates were 18/40 (45%) vs. 12/37 (32%) for abatacept and placebo treated patients.

Salivary gland item

For the salivary gland item, Based on expert opinion, unstimulated whole saliva secretion (UWS), stimulated whole saliva secretion (SWS), and salivary gland ultrasonography (SGUS) were selected for analysis. For UWS and SWS only relative change was explored because of wide inter-patient variation in secretion rates. ROC analysis for UWS showed a cut-off point of 0·0% (appendix p 11Table 3)., which Since this cut-off point falls within measurement variation., Therefore, a stricter cut-off point of ≥25% increase was selected.[20] ROC analysis for SWS showed a cut-off point of +44·2%. In line with UWS, a cut-off point of ≥25% increase was also analysed. For patients with no saliva production (0 ml/min) at baseline, any increase in saliva secretion was defined as response.

Since UWS showed higher response rates in rituximab trials (Supplementary Table 6appendix p 9), and is more commonly applied in pSS patients and more uniform and feasible to perform than SWS, UWS was included in the CRESS.

ROC analyses for absolute and relative change in total Hocevar score (SGUS) showed somewhat better discrimination for relative change, with a cut-off point of -14·7% (appendix p 11Table 3).[17][21] To minimize response due to natural variation, the cut-off point was set at ≥25% decrease in total Hocevar score. For additional SGUS analyses see Supplementary Table 7appendix p 13-14.

Since UWS and SGUS are complementary tests, measuring gland function and structural changes in the glands, both were selected. When combining these measurements, response rates were 23/40 (58%) vs. 15/37 (41%) for abatacept and placebo treated patients (Table 24).

Serological item

For the serological item, lLevels of RF, Immunoglobulin G (IgG), complement (C3 and C4), and lymphocyte count were considered as possible measurements by the experts. Only Rrelative change in serological markers was explored because of wide inter-patient variation in these scores. ROC analysis for RF levels showed a cut-off point of -23·0%, with good discrimination between treatment groups. (Table 3appendix p 11). Therefore, the cut-off point was set at ≥25% decrease. ROC analysis for IgG showed a cut-off point of -2·2% (appendix p 11). Because of the relatively small effect on IgG of abatacept (Table 24), and different

biological drugs might induce different serological responses, IgG response was also evaluated in patients on rituximab (~~Supplementary Table 6~~appendix p 9). The cut-off point of ≥10% was selected, giving acceptable sensitivity and specificity. ~~Analyses of S~~serum complement and lymphocyte count had no additional value (appendix p 15).~~are described in the Supplementary material. Both parameters were of no additional value.~~

~~When c~~Combining the complementary serological markers, RF and/or IgG response rates were 25/40 (63%) vs. 7/37 (19%) for abatacept and placebo treated patients (Table ~~2~~4).

~~*Definition of total CRESS response*~~

For a visual overview of CRESS items see Figure 1A. Response on ≥3 out of 5 items was most discriminating between treatment groups and therefore selected as definition of total CRESS response (Table ~~3~~5). Total CRESS response rates were 24/40 (60%) vs. 7/39 (18%) for abatacept and placebo treated patients (p<0·0001) (Table ~~2~~4, Figure 1B).

~~*Definition of total concise (c)CRESS response*~~

If OSS and/or SGUS are not available, the concise CRESS (cCRESS) can be used, leaving Schirmer's test and UWS for assessment of tear and salivary gland items, respectively. Total cCRESS response rates were 25/40 (63%) vs. 3/39 (8%) for abatacept and placebo treated patients (p<0·0001) (~~Table 6~~appendix p 16).

~~**Balance of CRESS items**~~

The number of CRESS responders who are responder on single items was well-balanced for all items (Figure 1C and appendix p 17). ~~When removing one of the five CRESS items and assessing responders for the remaining four items, independently of which item is excluded, discrimination remains similar: for responders on ≥2/4 and ≥3/4 items, there are equal differences in response percentages between active treatment and placebo groups (Table 7).~~

Criterion validity was confirmed by exploring the agreement of total CRESS response with PhyGDA (~~Supplementary Table~~appendix p ~~18~~8).

~~**External validation**~~

For external validation, ~~In TRACTISS and ETAP trials,~~ the cCRESS was analysed in TRACTISS and ETAP trials, since OSS and SGUS Hocevar score were not available. The serological item in the ETAP trial was based on IgG only, since numeric RF values were not available. Furthermore, IgG was only measured at week 12 and ~~was~~ imputed at week 24. In the multinational abatacept trial, the CRESS without SGUS was analysed.

Total CRESS or cCRESS response rates at the different time points are shown in Figure 2. Validation of the (c)CRESS in the TRACTISS and multinational abatacept trial as potentially positive trials demonstrated significant discrimination at the primary endpoint visit of the TRACTISS trial: ~~At the primary endpoint visit, (c)CRESS response rates were in the TRACTISS trial:~~ 33/67 (49%) rituximab vs. 20/66 (30%) for placebo (p=0·026), and a trend towards significance for the ~~in the~~ multinational abatacept trial: 41/92 (45%) for abatacept vs. 30/95 (32%) for placebo (p=0·067). The ETAP trial was identified as a negative trial, which cCRESS confirmed with low response rates in both treatment groups: ~~and in the ETAP trial:~~ 10/55 (18%) for tocilizumab vs. 13/55 (24%) for placebo (p=0·482).

~~Validation of the (c)CRESS in the TRACTISS and multinational abatacept trial as potentially positive trials demonstrated significant discrimination at the primary endpoint visit. For the multinational abatacept trial, a trend towards significance was found. The ETAP trial was identified as a negative trial, which cCRESS confirmed with low response rates in both treatment groups.~~

As shown in Figure 3, CRESS resulted in low response rates of placebo treated patients. In the multinational abatacept and ETAP trials, placebo response rates were 24-32% when using CRESS, compared to 58-64% when using ~~–~~ ESSDAI MCII ~~of~~ (Δ≥3) ~~points~~. In the TRACTISS trial, CRESS was able to demonstrate a higher response rate for rituximab compared to placebo, whereas no distinction could be made based on ESSDAI MCII response between both treatment groups at ~~the~~ primary endpoint visit. See appendix p 20-25~~Supplementary Tables 10-15~~ for external validation results of separate CRESS items and appendix p 26-27 for supportive analyses of selected CRESS items.

~~**Supportive analyses of selected CRESS items**~~

~~To support our selection of ClinESSDAI low disease activity (<5), ClinESSDAI<5 was compared to ClinESSDAI MCII (Δ≥3) and (c)CRESS was compared to (c)CRESS using the ClinESSDAIΔ≥3 instead of ClinESSDAI<5 in all four RCTs. Especially for the TRACTISS trial, total cCRESS using ClinESSDAI<5 showed better discrimination when compared to total cCRESS using ClinESSDAIΔ≥3 (Supplementary Table 16).~~

~~Finally, (c)CRESS was compared to only ClinESSDAI and ESSPRI response. Across all four RCTs, response rates on ClinESSDAI and ESSPRI showed no discrimination or less discrimination when compared to the (c)CRESS (Supplementary Table 17).~~

DISCUSSION

The choice of a primary study endpoint is a crucial step in designing clinical trials. Due to the heterogeneity and complexity of pSS, a composite endpoint is needed to reflect all important aspects of pSS. In the present study, we developed and validated the CRESS, a feasible composite endpoint combining disease activity, functional, and serological parameters, which enables discrimination between active treatment and placebo.

Our multidisciplinary expert team selected five complementary, clinically relevant items for the CRESS: systemic disease activity, patient-reported symptoms, tear gland, salivary gland, and serology. We found that all five CRESS items contributed equally to total CRESS response in the ASAP-III trial. When evaluating disease activity, systemic and patient-reported symptoms are both important. The validated instrument ESSDAI is widely used by physicians and researchers to assess systemic disease activity, and ESSPRI by patients to report prominent symptoms. High systemic disease activity may lead to severe complications of pSS, while patient-reported symptoms have greater impact on quality of life (QoL) of pSS patients.[2,21,32] Previous studies have demonstrated that ~~the~~ correlation between ESSDAI and ESSPRI is poor.[24,33] This indicates that these indices are complementary to each other, which is the reason to include both in the CRESS. ClinESSDAI low disease activity was selected for the systemic disease activity item, since this was more discriminating and might be more relevant for patients than the ClinESSDAI MCII (Δ≥3). ~~a decrease of the MCII of ≥3 points in ClinESSDAI, especially in patients with high disease activity at baseline. Low disease activity was more discriminating between treatment groups than the MCII. Furthermore, multiple previous trials failed to demonstrate differences between active treatment and placebo groups when using ESSDAI MCII as primary endpoint.~~ Low disease activity as response criterion has also been used in systemic lupus erythematosus ~~(SLE)~~. The Lupus Low Disease Activity State (LLDAS) has been developed and validated in several RCTs[25,34] and was found to be associated with reduced damage accrual and higher health-related ~~quality of life~~QoL.[25,26,27] The ESSPRI, consisting of ~~only~~ three questions regarding dryness, fatigue and pain, was preferred to more elaborate questionnaires assessing separate patient-reported features, ~~such as the Multidimensional Fatigue Inventory (MFI) and Xerostomia Index (XI),~~ which might be more appropriate to use as secondary endpoints in trials.

For the tear and salivary gland items, two complementary tests assessing glandular function and structural abnormalities were included: Schirmer's test and OSS for the tear gland ~~item~~ and UWS and SGUS for the salivary gland ~~item~~.[15] For Schirmer's test, variability is high, especially within normal scores.[28,37] Therefore, improvement in Schirmer's test was considered clinically relevant only for patients with abnormal values, thus definition of response was split for normal and abnormal values at baseline. ~~For the salivary gland item,~~

Formatted: Font: (Default) Arial

UWS was preferred to SWS~~,~~ because UWS is also included as an ACR-EULAR classification criterion[19] and is more uniform and feasible to perform. In contrast, SWS ~~can be~~is performed using gustatory (~~such as~~e.g. citric acid) or mechanical stimulation (chewing). SGUS was included because it is an up-coming, non-invasive tool for imaging, revealing structural changes of salivary glands. Because it is not yet clear which items of SGUS scores are most important in evaluation of therapeutic response and what each of these items reflect, we selected the widely applied total Hocevar score for the CRESS. Previously, we found~~showed~~ that reliability varies between ~~the~~ individual items of the Hocevar score. Total Hocevar score ~~also~~ showed~~s~~ excellent intra-observer and good inter-observer reliability, although total scores varied more between observers in patients with higher ultrasonographic scores.[28~~9~~] Therefore, relative change in total Hocevar score was considered as most applicable, and the cut-off selected was ~~set at~~ ≥25%. ~~Preferably, each patient should be scored by the same ultrasonographer at every time point. Our additional analyses indicated that the homogeneity item of the Hocevar score was most discriminating between abatacept and placebo groups. Although homogeneity shows high correlation with hypoechogenic areas[29], in our ASAP-III study hypoechogenic areas did not discriminate between treatment groups. Two previous RCTs found significant improvement in ultrasonography scores after rituximab treatment, with one study showing effect based on a combination of homogeneity and hypoechogenic areas.[30] The other study found significant improvement in total ultrasound score and glandular border definition, but not in hypoechogenic areas or homogeneity.[31] The recent OMERACT initiative demonstrates good reliability of a novel scoring system in which homogeneity and anechoic or hypoechoic areas play a central role.[32]~~ Further research and international collaboration is needed to determine the optimal SGUS scoring system for assessment of therapeutic response.

We decided to include the ClinESSDAI and serology as separate items~~,~~ because the biological domain of the ESSDAI is too global, based on cut-offs for IgG, meaning that improvement in IgG can be undetected. Moreover, the biological domain does not include RF autoantibodies, whereas RF is an important marker for disease activity and has been associated with development of MALT.[30~~3~~,34] ~~IgM-RF has a short half-life and is mostly produced by short-lived plasma cells, implying that an effect of immunotherapy affecting their precursors can be observed within weeks.[35]~~ RF may be evaluated ~~using either~~as total RF or IgM-RF, depending on the assay method. The vast majority of total RF consists of IgM-RF and response on this serological item is defined using relative change. Therefore, the cut-off point of ≥25% can be applied among different assays~~,~~ as was also the case in the validation trials. Immunologic and kinetic differences between formation of IgM-RF and IgG provide a rationale for combined analysis in clinical trials~~;~~ IgM-RF (shorter half-life) is mainly produced by short-lived plasma cells[31] and IgG antibodies (longer half-life) are mainly produced by

long-lived plasma cells. This is illustrated by the modest effect of B-cell depletion therapy on IgG levels and persistence of Ig-producing cells in salivary gland tissue.[9,32] ~~A significant decrease in IgG may take months, since IgG antibodies have a longer half-life and their production is more dependent on long-lived plasma cells. This is illustrated by the modest effect of B-cell depletion therapy on IgG levels and persistence of Ig-producing cells in salivary gland tissue.[9,36] Immunologic and kinetic differences between formation of RF and IgG provide a rationale for combined analysis in clinical trial setting. Although SSA auto-antibodies are an important serological abnormality in pSS patients, anti-SSA levels are not a feasible quantifiable measurement for all centres performing clinical trials.~~

~~We found that all five CRESS items contributed equally to total CRESS response in the ASAP-III trial. Four patients (two abatacept, two placebo) were responder on only the tear gland, salivary gland and serological item. All other CRESS responders were responder on either the systemic disease activity and/or the patient-reported symptoms item, in combination with any of the other items. The balanced combination of five items enables inclusion of a broader selection of patients. Also, no significant differences were found in baseline characteristics of CRESS responders versus non-responders (data not shown), indicating no major selection bias in CRESS responders. Recent RCTs which used the ESSDAI as primary endpoint, included patients with moderate to high systemic disease activity (ESSDAI≥5). However, this leaves out a large group of pSS patients, especially those with sicca complaints and fatigue, but no systemic involvement. With the CRESS as primary endpoint, these patients can also be included in clinical trials.~~

In the two external validation RCTs that were identified as potentially positive trials, ~~the TRACTISS and multinational abatacept trial,~~ significant discrimination with cCRESS was seen at the primary endpoint visit in the TRACTISS trial (week 48: 49% vs. 30%), and a trend towards significance with CRESS was seen in the multinational abatacept trial (week 24: 45% vs. 32%). The ETAP trial was identified as negative trial, which cCRESS confirmed, with low response rates in both treatment groups (week 24: 18% vs. 24%).

In the external validation trials ~~in which~~with high baseline ESSDAI ~~scores were applied~~ as ~~an~~ inclusion criterion (multinational abatacept and ETAP), CRESS was able to approximately half placebo response rates compared to ESSDAI MCII ~~response in the placebo groups was~~ (24-32% vs. 58-64%). ~~, whereas ESSDAI MCII response was 58-64%. CRESS was able to approximately half placebo response rates compared to ESSDAI MCII.~~ Patients with~~out not~~ enough items available to evaluate CRESS response were imputed as non-responder. Even w~~W~~hen excluding patients with <3 CRESS items available, discrimination between treatment groups remained similar, ~~when analysing CRESS response in all patients~~ and placebo

response in all trials ~~was still~~remained low (27-354%). This is an important improvement, since limited placebo response is essential to demonstrate treatment efficacy.

In the TRACTISS trial, no early treatment effect was seen at week 24. A longer follow-up period of 48 weeks with repeated cycles of rituximab might be needed to establish a treatment effect. The~~A~~ relatively high placebo response ~~was seen~~ at week 24. ~~This~~ may ~~partly~~ be ~~related to~~explained by the high dose of glucocorticoids. When analysing the individual CRESS items, there was a high response for ClinESSDAI <5 in both treatment groups (week 48: 76% vs. 60%). This can be explained because TRACTISS did not apply~~use~~ moderate to high systemic disease activity at baseline as ~~an~~ inclusion criterion, as opposed to the other trials. However, the alternative of assessing the (Clin)ESSDAI MCII leads to low response rates, since for patients with already low ESSDAI scores it would be impossible to improve, which restricts ~~. This would effectively lead a strict total CRESS response criterion of~~ response to~~on~~ ≥3/4 items. As seen from our additional analyses, replacing ClinESSDAI low disease activity by ClinESSDAI MCII in the cCRESS led to less discrimination between treatment groups in the TRACTISS trial. From a patients' point of view, it is also important that drugs become available for patients with a high symptom burden and a relatively low systemic disease activity, which is often the case in newly diagnosed pSS patients. All recent RCTs which used the ESSDAI as primary endpoint, included patients with moderate to high systemic disease activity (ESSDAI≥5). With the CRESS, patients with a relatively low systemic disease activity can also be included in clinical trials.

~~The importance of using a composite endpoint in pSS is confirmed by the use of composite measures in other heterogeneous, immune-mediated diseases. In myositis, core set measures for disease activity have been developed and validated for use in clinical studies and therapeutic trials.[37] Furthermore, in diffuse cutaneous systemic sclerosis, the revised Composite Response Index in Systemic Sclerosis (CRISS) has been developed, including five core set measures.[38] For the CRESS, we used different clinically relevant cut-off points for improvement per item instead of a fixed percentage of improvement for all items, because the clinical interpretation and measurement scales of the five CRESS items are not comparable.~~

Limitations to this study are that data on sensitivity to change of the CRESS in pSS patients with long-standing disease is not yet available, ~~,~~since RCTs included patients with a relatively short disease duration. ~~The RCTs in which the CRESS was developed and validated focussed on patients with a relatively short disease duration. Therefore, data on the sensitivity to change of the CRESS in pSS patients with long-standing disease is not yet available.~~ If patients with long-standing disease and ~~some degree of~~ damaged ~~in the~~ glands are included, certain CRESS items ~~of the CRESS~~ might be less sensitive to change,

especially the tear and salivary gland items, and the dryness component of the ESSPRI. However, objectively measured glandular function and sicca symptoms are not always coupled. Furthermore, it is ~~still~~ unclear whether dryness can improve ~~when~~ in damaged glands ~~are damaged~~. Since symptoms due to damage are not scored in the ClinESSDAI, this item and the serological item are~~is~~ expected to remain sensitive to change. ~~The same is expected for the serological item.~~

The CRESS was developed foremost based on expert opinion. Another composite endpoint, the Sjögren's Syndrome Responder Index (SSRI), includes five domains (visual analogue score for fatigue, ocular and oral dryness, UWS, and ESR) and was developed in 2015 based on a more data-driven approach using data of rituximab trials.[3~~9~~] ~~However, this~~This endpoint lacks important measures for pSS such as systemic disease activity and objective measures for tear gland function. Moreover, the SSRI response was inferior to CRESS response in the ASAP-III trial, with response rates at week 24 of 13/40 (33%) in the abatacept and 9/39 (23%) in the placebo group.[5] Currently, an Innovative Medicines Initiative (IMI) project, NECESSITY[3~~4~~~~0~~], is working on finding a new composite endpoint based on a data-driven approach of candidate criteria and a Delphi procedure, in which the CRESS will also be included for further analyses.

In conclusion, the CRESS is a feasible composite endpoint which consists of well-balanced, complementary, and clinically relevant items for all pSS patients. We showed that the CRESS is able to discriminate between active treatment and placebo in pSS patients and thus demonstrate clinical efficacy. As next step, validation of the CRESS in a future, prospective RCT is warranted. This study is an important advance in the search for valid composite endpoints for use in future clinical trials in pSS.

STATEMENTS

Competing interests statement

S. Arends: no competing interests

L. de Wolff: no competing interests

J.F. van Nimwegen: speaker and consultant for Bristol Myers Squibb

G.M. Verstappen: no competing interests

J. Vehof: ~~no competing interests~~speaker and consultant for Santen and Tramedico

M. Bombardieri: unrestricted grants from Amgen/Medimmune and Janssen, consultant for Amgen/Medimmune, Janssen, GSK, UCB.

S.J. Bowman: receives part funding of his salary from the NIHR Birmingham Biomedical Research Centre, consultant for Novartis, Abbvie and Galapagos and Abbvie

E. Pontarini: no competing interests

A.N. Baer: consultant for Bristol-Myers Squibb~~, Novartis and Abbvie~~ and Sanofi-Aventis

J. Gottenberg: received honoraries and research grants from BMS and Pfizer and honoraries from CSL Behring, Lilly, Janssen, UCB, Roche

R. Felten: no competing interests

M. Nys: is an employee and shareholder of Bristol Myers Squibb

N. Ray: is an employee and shareholder of Bristol Myers Squibb

A. Vissink: unrestricted grants from Bristol Myers Squibb and Roche

F.G.M. Kroese: unrestricted grants from Bristol Myers Squibb, consultant for Bristol Myers Squibb, speaker for Bristol Myers Squibb, Roche and Janssen-Cilag

H. Bootsma: unrestricted grants from Bristol Myers Squibb and Roche, consultant for Bristol Myers Squibb, Roche, Novartis, Medimmune, Union Chimique Belge, speaker for Bristol Myers Squibb and Novartis

Contributors

SA, JV, GMV, AV, NR, FGMK, HB were part of the multidisciplinary expert team. SA and LW performed the statistical analysis for the development and validation, EP performed the statistical analysis for the validation in the TRACTISS trial, MN performed the statistical analysis for the validation in the multinational abatacept trial. SA and LW wrote the first draft of the article. All authors interpreted the data, contributed to writing, critically revised the article, and approved the final submitted version.

Acknowledgements

The authors would like to acknowledge valuable discussions with M. Nowak, D. Grasela and S. Banerjee of BMS and with K. Hammitt of the Sjögren's Foundation.

Trial contributors

~~The authors would like to acknowledge the different trial contributors.~~ The ASAP-III trial: E. Mossel, G. van Zuiden, K. Delli, A. Stel, B. van der Vegt, E. Haacke, L. Olie, L. Los, S. Pringle and F. Spijkervet; the TRACTISS trial: Leeds CTRU and the colleagues involved in the TRACTISS trial: C. Everett, J. O'Dwyer, P. Emery, C. Pitzalis, W. Ng, C. Pease, E. Price, N. Sutcliffe, N. Gendi, F. Hall, S. Ruddock, C. Fernandez, C. Reynolds, C. Hulme, K. Davies, C. Edwards, P. Lanyon, R. Moots, E. Roussou, I. Giles, L. Sharples~~, which was funded by Versus Arthritis~~; the multinational abatacept trial: E. St Clair, T. Sumida, T. Takeuchi, G. Foulks, S. Mukherjee, R. Wong, B. Gavin; and the ETAP trial: V. Devauchelle-Pensec, R. Seror, P. Duffau, D. Saadoun, E. Hachulla, PY. Hatron, C. Salliot, A. Perdriger, J. Morel, A. Mékinian, O. Vittecoq, JM. Berthelot, E. Dernis, V. Le Guern, P. Dieudé, C. Larroche, C. Richez, T. Martin, C. Zarnitsky, G. Blaison, P. Kieffer, F. Maurier, A. Dellal, S. Rist, E. Andres, A Contis, E. Chatelus, C. Sordet, J. Sibilia, C. Arnold, M. Tawk, O. Aberkane, L. Holterbach, P. Cacoub, A. Saraux, X. Mariette and N. Meyer.

Data sharing

Currently there are no plans to share additional data beyond what is shared in this article.

REFERENCES

1. Mariette X, Criswell LA. Primary Sjögren's syndrome. N Engl J Med. 2018;378(10):931–9.
2. Ramos-Casals M, Brito-Zerón P, Bombardieri S et al. EULAR recommendations for the management of Sjögren's syndrome with topical and systemic therapies. Ann Rheum Dis. 2020;79(1):3–18.
3. Meiners PM, Vissink A, Kroese FGM et al. Abatacept treatment reduces disease activity in early primary Sjögren's syndrome (open-label proof of concept ASAP study). Ann Rheum Dis. 2014;73(7):1393–6.
4. Meiners PM, Arends S, Brouwer E, Spijkervet FKL, Vissink A, Bootsma H. Responsiveness of disease activity indices ESSPRI and ESSDAI in patients with primary Sjögren's syndrome treated with rituximab. Ann Rheum Dis. 2012;71(8):1297–302.
5. van Nimwegen JF, Mossel E, van Zuiden GS et al. Abatacept treatment for patients with early active primary Sjögren's syndrome: a single-centre, randomised, double-blind, placebo-controlled, phase 3 trial (ASAP-III study). Lancet Rheumatol. 2020;9913(19):1–11.
6. Baer A, Gottenberg J-E, St.Clair EW et al. Efficacy and safety of abatacept in active primary Sjögren's syndrome: results of a phase III, randomised, placebo-controlled trial. Ann Rheum Dis. 2020;0:1–10.
7. Felten R, Devauchelle-Pensec V, Seror R et al. Interleukin 6 receptor inhibition in primary Sjögren syndrome: A multicentre double-blind randomised placebo-controlled trial. Ann Rheum Dis. 2020;0:1–10.
8. Bowman SJ, Everett CC, O'Dwyer JL et al. Randomized Controlled Trial of Rituximab and Cost-Effectiveness Analysis in Treating Fatigue and Oral Dryness in Primary Sjögren's Syndrome. Arthritis Rheumatol. 2017;69(7):1440–50.
9. Devauchelle-Pensec V, Mariette X, Jousse-Joulin S et al. Treatment of primary sjögren syndrome with rituximab a randomized trial. Ann Intern Med. 2014;29(1):6–12.
10. de Wolff L, Arends S, van Nimwegen JF, Bootsma H. Ten years of the ESSDAI: is it fit for purpose? Clin Exp Rheumatol. 2020;38(4):283–90.
11. Rider L.G., Aggarwal R., Machado P.M.,et al. Update on outcome assessment in myositis. Nat Rev Rheumatol. 2018 ;14(5) :303-18.
12. Khanna D, Huang S, Lin CJF, Spino C. New composite endpoint in early diffuse cutaneous systemic sclerosis: revisiting the provisional American College of Rheumatology Composite Response Index in Systemic Sclerosis. Ann Rheum Dis. 2020;0:1–10.

13. Meijer JM, Meiners PM, Vissink A et al. Effectiveness of rituximab treatment in primary sjögren's syndrome: A randomized, double-blind, placebo-controlled trial. Arthritis Rheum. 2010;62(4):960–8.

14. Seror R, Bootsma H, Saraux A et al. Defining disease activity states and clinically meaningful improvement in primary Sjögren's syndrome with EULAR primary Sjögren's syndrome disease activity (ESSDAI) and patient-reported indexes (ESSPRI). Ann Rheum Dis. 2016;75(2):382–9.

15. Seror R, Meiners P, Baron G et al. Development of the ClinESSDAI: A clinical score without biological domain. A tool for biological studies. Ann Rheum Dis. 2016;75(11):1945–50.

16. Wolffsohn JS, Arita R, Chalmers R et al. TFOS DEWS II Diagnostic Methodology report. Ocul Surf. 2017;15(3):539–74.

17. Whitcher JP, Shiboski CH, Shiboski SC et al. A Simplified Quantitative Method for Assessing Keratoconjunctivitis Sicca From the Sjögren's Syndrome International Registry. Am J Ophthalmol. 2010;149:405-415.

~~16. Ericsson Y, Hardwick L. Individual diagnosis, prognosis and counselling for caries prevention. Caries Res. 1978;12(Suppl 1):94-102.~~

~~17. Hočevar A, Rainer S, Rozman B, Zor P, Tomšič M. Ultrasonographic changes of major salivary glands in primary Sjögren's syndrome. Evaluation of a novel scoring system. Eur J Radiol. 2007;63(3):379–83.~~

18. Boujnah Y, Mouchel R, El-Chehab H, Dot C, Burillon C, Kocaba V. Prospective, monocentric, uncontrolled study of efficacy, tolerance and adherence of cyclosporin 0.1 % for severe dry eye syndrome. J Fr Ophtalmol. 2018;41(2):129–35.

19. Shiboski CH, Shiboski SC, Seror R et al. 2016 American College of Rheumatology/European League Against Rheumatism classification criteria for primary Sjögren's syndrome: A consensus and data-driven methodology involving three international patient cohorts. Ann Rheum Dis. 2017;76(1):9–16.

20. Burlage FR, Pijpe J, Coppes RP et al. Variability of flow rate when collecting stimulated human parotid saliva. Eur J Oral Sci. 2005;113(5):386–90.

21. Hočevar A, Rainer S, Rozman B, Zor P, Tomšič M. Ultrasonographic changes of major salivary glands in primary Sjögren's syndrome. Evaluation of a novel scoring system. Eur J Radiol. 2007;63(3):379–83.

22. Lendrem D, Mitchell S, McMeekin P et al. Health-related utility values of patients with primary Sjögren's syndrome and its predictors. Ann Rheum Dis. 2014;73(7):1362–8.

23. Hammitt KM. Sjögren's: the patients' perspective. Clin Exp Rheumatol 2020;38(Suppl 126)25–6.

24. Seror R, Theander E, Brun JG et al. Validation of EULAR primary Sjögren's syndrome

Formatted: English (United Kingdom)

Formatted: English (United Kingdom)

disease activity (ESSDAI) and Patient indexes (ESSPRI). Ann Rheum Dis. 2015;74(5):859–66.

25. Morand EF, Trasieva T, Berglind A, Illei GG, Tummala R. Lupus Low Disease Activity State (LLDAS) attainment discriminates responders in a systemic lupus erythematosus trial: Post-hoc analysis of the Phase IIb MUSE trial of anifrolumab. Ann Rheum Dis. 2018;77(5):706–13.

26. Tsang-A-Sjoe MWP, Bultink IEM, Heslinga M, Voskuyl AE. Both prolonged remission and Lupus Low Disease Activity State are associated with reduced damage accrual in systemic lupus erythematosus. Rheumatology (Oxford). 2017;56(1):121–8.

27. Golder V, Kandane-Rathnayake R, Hoi AYB et al. Association of the lupus low disease activity state (LLDAS) with health-related quality of life in a multinational prospective study. Arthritis Res Ther. 2017;19(1):1–11.

28. Nichols KK, Mitchell GL, Zadnik K. The Repeatability of Clinical Measurements of Dry Eye. Cornea. 2004;23(3):272–85.

29. Delli K, Arends S, van Nimwegen JF et al. Ultrasound of the Major Salivary Glands is a Reliable Imaging Technique in Patients with Clinically Suspected Primary Sjögren's Syndrome. Ultraschall Med 2018;39:328-33.

~~29. Mossel E, Arends S, Van Nimwegen JF et al. Scoring hypoechogenic areas in one parotid and one submandibular gland increases feasibility of ultrasound in primary Sjögren's syndrome. Ann Rheum Dis. 2018;77(4):556–62.~~

~~30. Jousse-Joulin S, Devauchelle-Pensec V, Cornec D et al. Brief Report: Ultrasonographic assessment of salivary gland response to rituximab in primary Sjögren's syndrome. Arthritis Rheumatol. 2015;67(6):1623–8.~~

~~31. Fisher BA, Everett CC, Rout J et al. Effect of rituximab on a salivary gland ultrasound score in primary Sjögren's syndrome: results of the TRACTISS randomised double-blind multicentre substudy. Ann Rheum Dis. 2018;77(3):412–6.~~

~~32. Finzel S, Jousse-Joulin S, Costantino F et al. Patient-based reliability of the Outcome Measures in Rheumatology (OMERACT) ultrasound scoring system for salivary gland assessment in patients with Sjögren's syndrome. Rheumatol. 2020;0:1-8.~~

30. Nocturne G, Virone A, Ng WF et al. Rheumatoid Factor and Disease Activity Are Independent Predictors of Lymphoma in Primary Sjögren's Syndrome. Arthritis Rheumatol. 2016;68(4):977–85.

~~34. James K, Chipeta C, Parker A et al. B-cell activity markers are associated with different disease activity domains in primary Sjögren's syndrome. Rheumatol (United Kingdom). 2018;57(7):1222–7.~~

31. William J, Euler C, Shlomchik MJ. Short-Lived Plasmablasts Dominate the Early Spontaneous Rheumatoid Factor Response: Differentiation Pathways, Hypermutating

Formatted: English (United Kingdom)

Cell Types, and Affinity Maturation Outside the Germinal Center. J Immunol. 2005;174(11):6879–87.

326. Hamza N, Bootsma H, Yuvaraj S et al. Persistence of immunoglobulin-producing cells in parotid salivary glands of patients with primary Sjögren's syndrome after B cell depletion therapy. Ann Rheum Dis. 2012;71(11):1881–7.

~~37. Rider L.G., Aggarwal R., Machado P.M.,et al. Update on outcome assessment in myositis. Nat Rev Rheumatol. 2018 ;14(5) :303-18.~~

~~38. Khanna D, Huang S, Lin CJF, Spino C. New composite endpoint in early diffuse cutaneous systemic sclerosis: revisiting the provisional American College of Rheumatology Composite Response Index in Systemic Sclerosis. Ann Rheum Dis. 2020;0:1–10.~~

339. Cornec D, Devauchelle-Pensec V, Mariette X et al. Development of the Sjögren's Syndrome Responder Index, a data-driven composite endpoint for assessing treatment efficacy. Rheumatol (United Kingdom). 2015;54(9):1699–708.

340. New clinical endpoints in primary Sjögren's Syndrome: an interventional trial based on stratifying patients (NECESSITY). Innovative Medicines Initiative. Grant agreement number 806975.

Formatted: French (France)

Tables and figure legends

Table 1: CRESS items and definition of response per item

Items	*Measurement*	*Definition of response*
Systemic disease activity	ClinESSDAI	Score<5 (low disease activity)
Patient-reported symptoms	ESSPRI	Decrease of ≥1 point or ≥15% from baseline
Tear gland*,**	Schirmer/OSS	If abnormal Schirmer (≤5 mm) at baseline: increase of ≥5 mm in Schirmer from baseline Or if abnormal OSS (≥3 points) at baseline: decrease of ≥2 points in OSS from baseline Or if both Schirmer and OSS normal score at baseline (Schirmer >5 mm and OSS <3 points): no change to abnormal in Schirmer and OSS
Salivary gland*	UWS/SGUS	Increase of ≥25%, or if score is 0 ml/min at baseline any increase in UWS from baseline Or decrease of ≥25% in total Hocevar score (SGUS) from baseline
Serological	RF***/IgG	Decrease of ≥25% in RF from baseline Or decrease of ≥10% in IgG from baseline
CRESS responder	**Responder on ≥3 of 5 items**	

Abbreviations: CRESS: Composite of Relevant Endpoints for Sjögren's Syndrome, ClinESSDAI: Clinical European League Against Rheumatism (EULAR) Sjögren's Syndrome Disease Activity Index, ESSPRI: EULAR Sjögren's Syndrome Patient Reported Index, Schirmer: Schirmer's test, OSS: Ocular Staining Score, UWS: Unstimulated whole saliva secretion, SGUS: Salivary Gland Ultransonography, RF: Rheumatoid factor, IgG: Immunoglobulin G

*The CRESS can be used without OSS and/or SGUS if not available (concise CRESS), leaving Schirmer and/or UWS for assessment of tear and salivary gland items (see Supplementary Table 5)

**Mean of both eyes

***Total RF or RF-IgM

Table 2: CRESS response rates at week 24 in patients treated with abatacept and placebo

Items	*Abatacept*	*Placebo*
Systemic disease activity		
ClinESSDAI <5 points week 24	**18/40 (45%)**	**10/37 (27%)**
Patient-reported symptoms		
ESSPRI decrease ≥1 point	*23/40 (58%)*	*8/36 (22%)*
ESSPRI decrease ≥15%	*21/40 (53%)*	*8/36 (22%)*
ESSPRI decrease≥1 point or 15%	**23/40 (58%)**	**8/36 (22%)**
Tear gland		
Schirmer increase ≥5 mm	*4/23 (17%)*	*0/23 (0%)*
OSS decrease ≥2 points	*12/26 (46%)*	*8/26 (31%)*
Stable >5 mm and <3 points	*3/5 (60%)*	*4/8 (50%)**
Schirmer increase ≥5 mm or OSS decrease ≥2 points or stable (>5 mm and <3 points)	**18/40 (45%)**	**12/37 (32%)**
Salivary gland		
*UWS increase ≥25%***	*14/29 (48%)*	*8/30 (27%)*
Any increase if BL=0	*3/8 (38%)*	*3/5 (60%)*
SGUS decrease ≥25%	*11/37 (30%)*	*4/35 (11%)*
UWS increase ≥25% or any increase if BL=0 or SGUS decrease ≥25%	**23/40 (58%)**	**15/37 (41%)**
Serological		
*RF decrease ≥25%****	*24/36 (67%)*	*3/34 (9%)*
IgG decrease ≥10%	*7/38 (18%)*	*5/37 (14%)*
RF decrease ≥25% or IgG ≥10%	**25/40 (63%)**	**7/37 (19%)**
Total CRESS response		
Responder ≥3 of 5 items	*24/40 (60%)*	*7/37 (19%)*
Responder ≥3 of 5 items†	**24/40 (60%)**	**7/39 (18%)**

Abbreviations: CRESS: Composite of Relevant Endpoints for Sjögren's Syndrome, ClinESSDAI: Clinical European League Against Rheumatism (EULAR) Sjögren's Syndrome Disease Activity Index, ESSPRI: EULAR Sjögren's Syndrome Patient Reported Index, Schirmer: Schirmer's test, OSS: Ocular Staining Score, UWS: Unstimulated whole saliva secretion, BL=baseline, SGUS: Salivary Gland Ultransonography, RF: Rheumatoid factor, IgG: Immunoglobulin G

*1 patient with missing Schirmer's test

**Value=0 at baseline excluded (n=13)

***Value=0 at baseline excluded (n=7)

†Imputing 2 placebo patients with missing data at week 24 as non-responder

Table 3: Abatacept and placebo patients who were responder on ≥1 to 5 items at week 24					
Responder on number of items	*≥1*	*≥2*	***≥3***	*≥4*	*5*
Abatacept (n=40)	39 (98%)	34 (85%)	**24 (60%)**	9 (23%)	1 (3%)
Placebo (n=37)	30 (81%)	15 (41%)	**7 (19%)**	0	0
Bold text indicates highest sum of sensitivity and specificity					

Figure legends

Figure 1 A. Visual overview of Composite of Relevant Endpoints for Sjögren's Syndrome (CRESS) items and total CRESS response, *Figure 1A was created with BioRender.com.* **B.** Percentage of CRESS responders at week 24 for abatacept and placebo treatment **C.** Number of CRESS responders who are responder or non-responder on the single items for abatacept and placebo treatment

Figure 2 A. CRESS and cCRESS (which is CRESS without OSS and SGUS) response rates in ASAP-III trial at week 12 and 24. At week 12 only CRESS without SGUS was available **B.** cCRESS response rates in TRACTISS trial at week 24 and 48 **C.** CRESS without SGUS response rates in multinational abatacept trial at week 12 and 24 **D.** cCRESS without RF response rates in ETAP trial at week 12 and 24

Figure 3 A. CRESS response rates vs. ESSDAI MCII response rates in ASAP-III trial at week 24. Median baseline ESSDAI scores for abatacept were 14·0 (IQR 9·0-16.8), for placebo 13·0 (IQR 8·0-18·0) **B.** cCRESS (which is CRESS without OSS and SGUS) response rates vs. ESSDAI MCII response rates in TRACTISS trial at week 48. Mean baseline ESSDAI scores for rituximab were 5·3, SD±4·7, for placebo 6·0, SD±4·3 **C.** CRESS without SGUS response rates vs. ESSDAI MCII response rates in multinational abatacept trial at week 24. Mean baseline ESSDAI scores for abatacept were 8·7, SD±3·4, for placebo 10·1, SD±5·0 **D.** cCRESS without RF response rates vs. ESSDAI MCII response rates in ETAP trial at week 24. Median baseline ESSDAI scores for tocilizumab were 11·0 (IQR 8·0-13·5), for placebo 10·0 (IQR 8·0-14·8). For all trials, ESSDAI response percentages were calculated imputing missing patients as non-responders.

Figure 1 (.pdf)

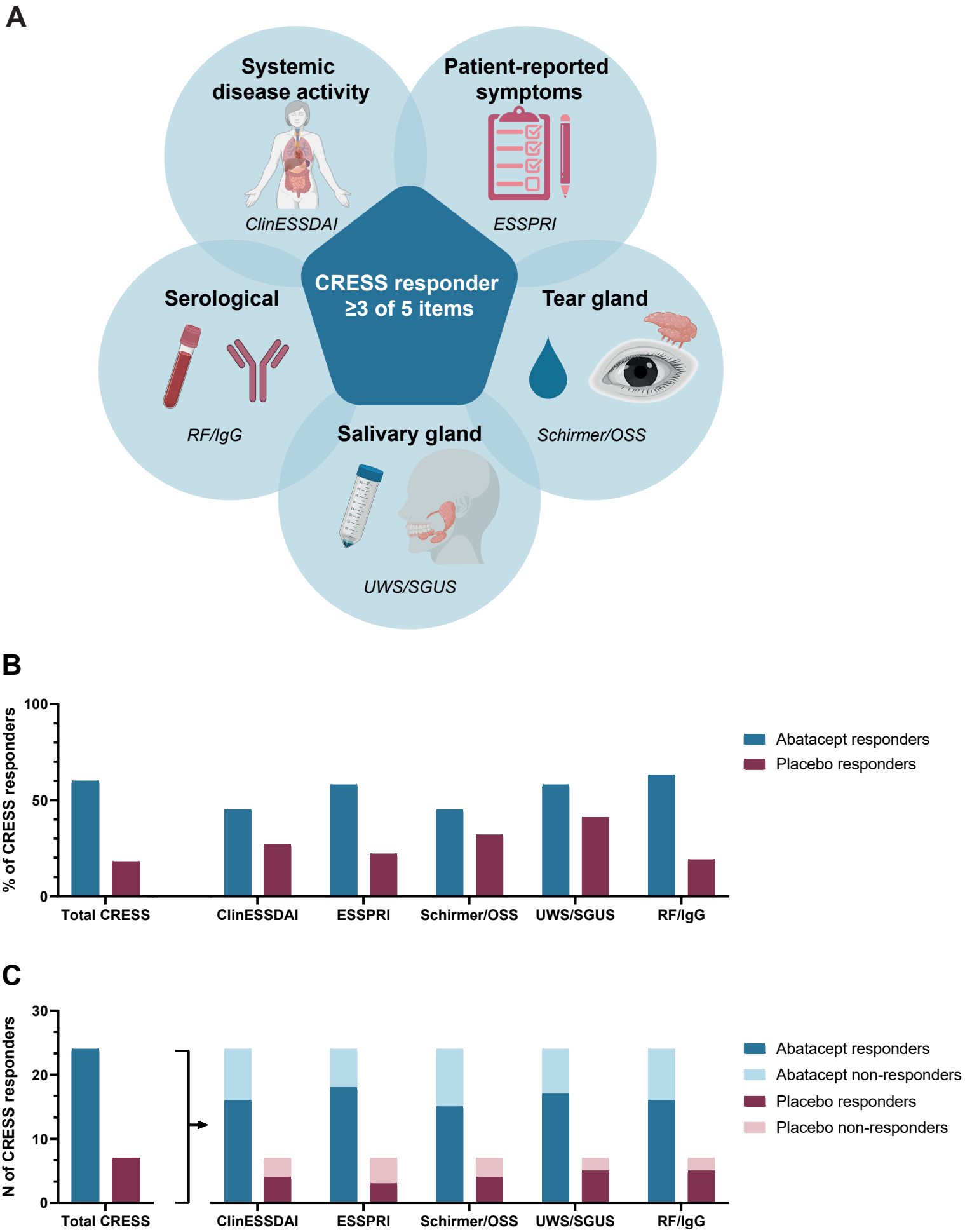

A
Systemic disease activity
ClinESSDAI
Patient-reported symptoms
ESSPRI
CRESS responder ≥3 of 5 items
Serological
RF/IgG
Tear gland
Schirmer/OSS
Salivary gland
UWS/SGUS
B
% of CRESS responders
100
50
0
Total CRESS
ClinESSDAI
ESSPRI
Schirmer/OSS
UWS/SGUS
RF/IgG
Abatacept responders
Placebo responders
C
N of CRESS responders
30
20
10
0
Total CRESS
ClinESSDAI
ESSPRI
Schirmer/OSS
UWS/SGUS
RF/IgG
Abatacept responders
Abatacept non-responders
Placebo responders
Placebo non-responders

Figure 2

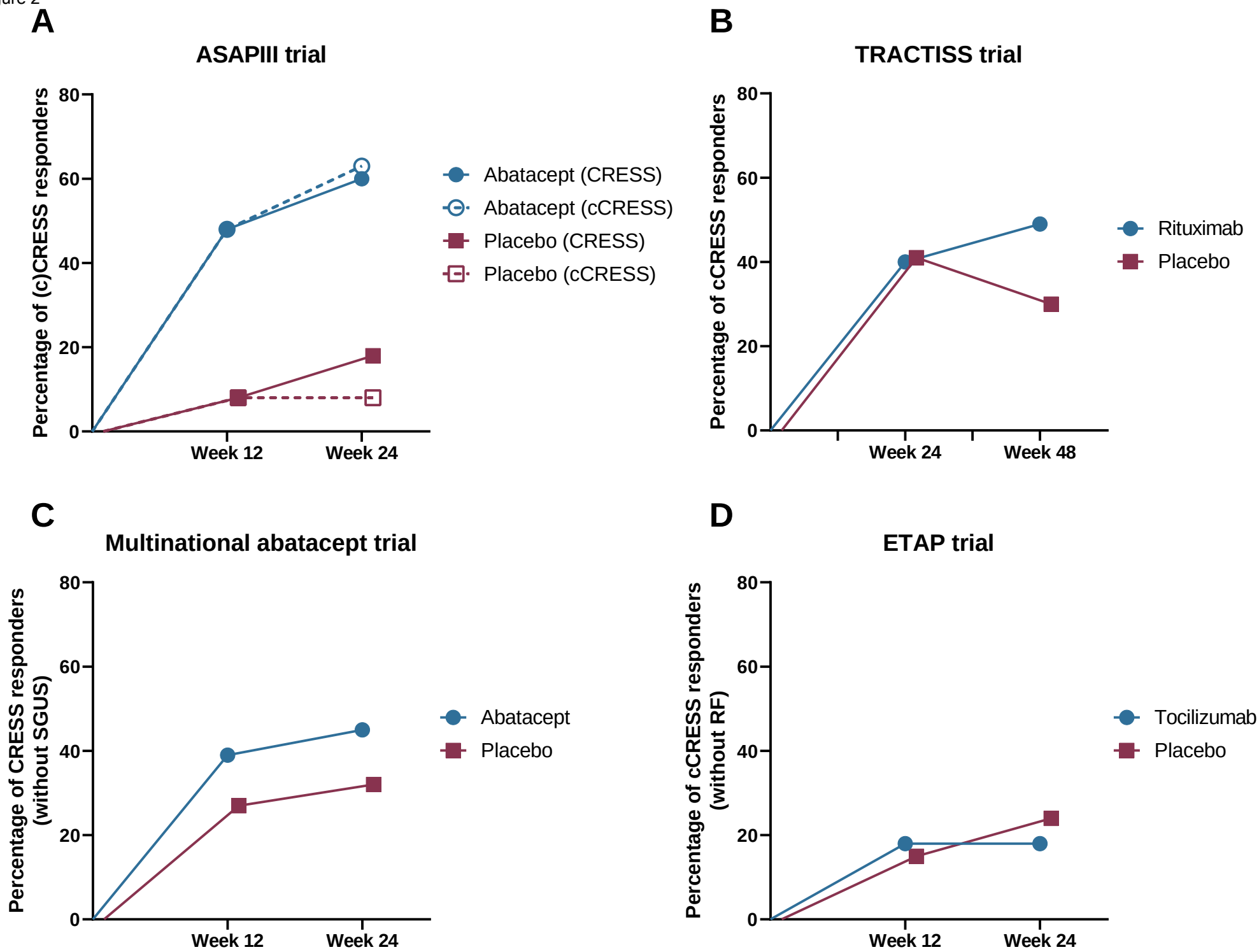
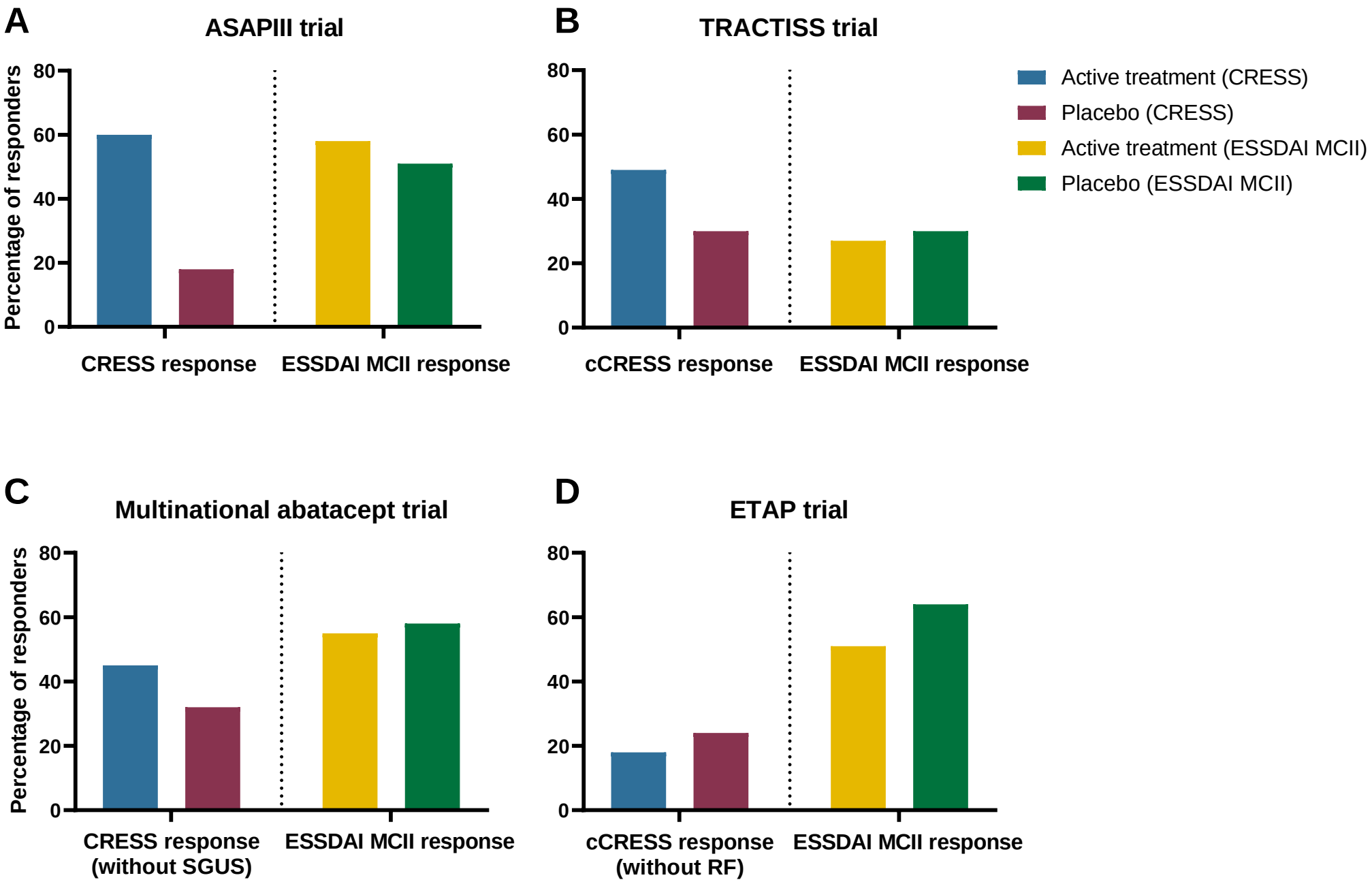
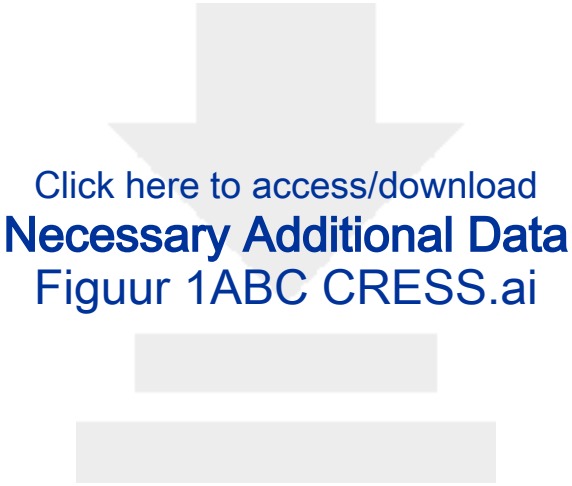

A
ASAPIII trial
Percentage of (c)CRESS responders
80
60
40
20
0
Week 12
Week 24
Abatacept (CRESS)
Abatacept (cCRESS)
Placebo (CRESS)
Placebo (cCRESS)
B
TRACTISS trial
Percentage of cCRESS responders
80
60
40
20
0
Week 24
Week 48
Rituximab
Placebo
C
Multinational abatacept trial
Percentage of CRESS responders (without SGUS)
80
60
40
20
0
Week 12
Week 24
Abatacept
Placebo
D
ETAP trial
Percentage of cCRESS responders (without RF)
80
60
40
20
0
Week 12
Week 24
Tocilizumab
Placebo

Figure 3


A
ASAPIII trial
Percentage of responders
80
60
40
20
0
CRESS response
ESSDAI MCII response
B
TRACTISS trial
cCRESS response
ESSDAI MCII response
Active treatment (CRESS)
Placebo (CRESS)
Active treatment (ESSDAI MCII)
Placebo (ESSDAI MCII)
C
Multinational abatacept trial
CRESS response (without SGUS)
D
ETAP trial
cCRESS response (without RF)

Appendix

Figure 1 (.ai)


Click here to access/download
Necessary Additional Data
Figuur 1ABC CRESS.ai

Editor's comments:

1. Title: The study title should have a descriptor—ie, randomised trial, case-control study, prospective analysis, population-based study etc—and should be non-declamatory. As such, please revise the title (eg, “Composite of Relevant Endpoints for Sjögren’s Syndrome (CRESS): development and validation of a novel outcome measure” or similar), keeping in mind that the title must be 180 characters or less (including spaces).

Answer: We changed the title.

2. Authorship: Please check with your co-authors, and confirm, that all names are spelled correctly, and affiliations are listed correctly. We cannot guarantee that we will be able to correct names and affiliations after publication of your article. Please also indicate which (if any) authors are full professors.

Answer: We checked the names of the co-authors and affiliations.

3. Authorship: Please format the author affiliation list to *Lancet* style. Please list authors by full first name and last name; and then for affiliations, by including the author initial and full last name, followed by one degree, in brackets following the author institution.

Answer: We formatted the author affiliation list to Lancet style.

4. Funding: The funding statement should pertain directly to this study (rather than to the original RCTs). If there was no direct funding for this study, please indicate this in the abstract. If the study did receive direct funding, please include a 'Role of the funding source' at the end of the methods. The following points should be specified in this section: (a) the role of the sponsors in the study design; (b) the role of the sponsors in the collection, analysis, or interpretation of the data; (c) the role of the sponsors in the writing of the report; (d) those who had access to the raw data (by author initials). If the funding source had no role then this should be stated (ie, "The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report"). Please also add to this section (if true, or amend if not): "The corresponding author had full access to all of the data and the final responsibility to submit for publication.

Answer: We added to the abstract that no direct funding was received for this study

5. Results: Please note that we do not allow subheadings in the results section, per Lancet style. Please reformat the section accordingly.

Answer: We removed the subheadings in the Results section.

6. Results: It is *Lancet* style to give actual numbers (numerator and denominator) together with all percentages, throughout the text and in tables etc. Please include n/N where percentages are given in the results section.

Answer: We added numbers together with the percentages.

7. Results: *Lancet* style is to provide p values to two significant figures, unless $p<0.0001$ (note number of decimal places). Please check that all p values are reported in this way.

Answer: We changed the p values accordingly.

8. Please include a formal discussion of the study limitations in the Discussion section.

Answer: We included a formal discussion of the study limitations in the Discussion section.

9. Your paper is substantially over our length and reference limits (3500 words, not including references, COI statements, abstract etc; maximum of 30 references). Please endeavour to reduce the length as much as possible; I can offer some leeway on this, but the text should not exceed 4000 words. Please also check the references carefully and remove any that are not essential.

Answer: We reduced the word count to 3994 words and reduced references from 40 to 34.

10. Only 5-6 non-text items (figures or tables; this also includes the Research in Context panel) can be accommodated in the print edition – you currently have 11 (7 tables, 3 figures, 1 panel). Please consider which of these items could be moved to the appendix such that the final manuscript includes a maximum of 7 display items.

Answer: We moved tables to the supplementary material so that the final manuscript includes a maximum of 7 display items (3 tables, 3 figures and 1 Research in Context panel).

11. Tables should be supplied in a separate Word file (not Excel or fdf/pdf). Each row of data should be in a separate line. Please ensure that rows and columns are not tabbed; data should be entered in cell form.

Answer: We uploaded the tables in a separate word file.

12. Please ensure that you provide your figures as individual, editable files. For statistical images such as histograms, survival or time-to-event curves, line graphs, scatter graphs, and forest plots you should provide editable vector files (ie, the original artwork generated by the statistical package used to make the image, typically by using "Export" or "Print to file" commands); our preferred formats for these files are .eps, .pdf, or .ai. Photographic images must be provided at a minimum of 300 dpi at 107 mm wide. We cannot guarantee accurate reproduction of images without these files. For more information, please see: https://www.thelancet.com/for-authors/forms?section=artwork

Answer: Checked.

13. Please submit an authors' contribution and signature form (attached here); both hand-written and electronic signatures are acceptable. Please ensure that the contributions listed on the form match those in the Author contributions section at the end of the paper.

Answer: We attached the forms to the submission.

14. Please provide completed ICMJE declaration forms from all authors (1 form per author) listed declaring any potential conflicts of interest (form attached here). Please ensure that the declarations listed on these forms exactly matches those in the "Declaration of interest" section at the end of the paper.

Answer: We attached the forms to the submission.

15. We need written confirmation from all individuals named in the Acknowledgments section to confirm that they are happy to be named in the paper. The following format can be used: "I permit Liseth de Wolff et al to list my name in the acknowledgments section of their manuscript and I have seen a copy of the paper, “Composite of Relevant Endpoints for Sjögren’s Syndrome (CRESS): a comprehensive tool for clinical trials [amended to the new title, as above]”. This can be provided by email and the compiled permissions uploaded as a single PDF file. Given that you have acknowledged a large number of individuals, you might consider listing the individuals from the previous trials in a separate section to avoid having to gather permissions from all of these people; however, if you wish to formally acknowledge them, we will need to have permission.

Answer: We attached the confirmations by email (in PDF) to the submission.

16. Please be sure to state whether this study was fully or in part NIH funded, if any authors are employed by NIH, or if any authors are in receipt of an NIH grant.

Answer: This study was not funded by NIH, no authors are employed by NIH or received NIH grants.

17. All submitted Articles must contain a data sharing statement, to be included at the end of the manuscript or in an appendix. Data sharing statements must indicate:

- Whether data collected for the study, including individual participant data and a data dictionary defining each field in the set, will be made available to others;
- What data will be made available (deidentified participant data, participant data with identifiers, data dictionary, or other specified data set);
- Whether additional, related documents will be available (eg, study protocol, statistical analysis plan, informed consent form);
- When these data will be available (beginning and end date, or "with publication", as applicable);
- Where the data will be made available (including complete URLs or email addresses if relevant);
- By what access criteria data will be shared (including with whom, for what types of analyses, by what mechanism - eg, with or without investigator support, after approval of a proposal, with a signed data access agreement - or any additional restrictions).
  See https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(17)31282-5/fulltext for examples. Clinical trials that begin enrolling participants on or after Jan 1, 2019, must include a data sharing plan in the trial's registration. If the data sharing plan changes after registration, this should be reflected in the statement submitted and published, and updated in the registry record. For reports of research other than clinical trials, data sharing statements are encouraged but not required. Mendeley Data (https://data.mendeley.com/) is a secure online repository for research data, permitting archiving of any file type and assigning a permanent and unique digital object identifier (DOI) so that the files can be easily referenced. If authors wish to share their supporting data, and have not already made alternative arrangements, a Mendeley DOI can be referred to in the data sharing statement.

Answer: We added a data sharing statement in the manuscript: Currently there are no plans to share additional data beyond what is shared in this article.

18. Appendix: Please supply the revised web appendix as a single PDF file, with the pages paginated. When you refer to an item in the appendix, please refer to the page number on which it appears, not the table or section (eg, appendix p 1).

Answer: We supplied the revised appendix as a single PDF file with paginated numbers and referred to the appendix page in the main manuscript.

19. It is not *Lancet* policy to edit or style supplementary material for the web; however, this material will be hosted on our website as a pdf conversion of the author-supplied file. Please style your supplementary material as per the guidelines below. Please note that we will be unable to correct any errors in the online appendix following publication; as such, please check carefully when submitting. Please supply the appendix as a single file, with numbered pages. Please do not include a cover page with details of the paper, as we add a cover page to the appendix file before publication with this information.
    Required formatting of data in the appendix:

- SI units are required, if appropriate
- Numbers in text and tables should always be provided if % is shown.
- Means should be accompanied by SDs, and medians by interquartile range.
- Exact p values should be provided, unless p<0·0001
- For drug names, recommended international nomenclature (rINN) is required
- References should be in Vancouver style. They should be numbered in order of mention in appendix and numbered separately from references in the full paper
- For figures in the appendix, all images must have a minimum resolution of 300 dpi at a width of 107 mm.

Answer: We checked formatting guidelines of the appendix and changed it accordingly.