Paper	Domain 1	Domain 2	Domain 3	Domain 4	Domain 5	Overall ROB
Efficacy and Safety of Upadacitinib in Patients With Moderate to Severe Atopic Dermatitis	1.1.: Stratified randomization was noted, but randomization technique used was not discussed (NI). 1.2: Double-blind was mentioned, but blinding was not discussed further. Clinical trial registration indicated quadruple masking, but no further information (PY) 1.3: No suspicious or excessive baseline imbalances or similarities noted (N) Overall: Some concerns	2.1: Study was double-blind and quadruple masked. Placebo was used (N). 2.2: Study was double-blind, and carers were masked. All placebo subjects were rerandomized to investigational treatment (PN). 2.6: ITT analysis was conducted, and is considered appropriate (Y) Overall: Low risk	3.1: Population for ITT analysis was randomized. Greater than 5% and less than 20% of patients were discontinued from the study (N). 3.2: Multiple imputation was conducted for both groups using the Markov Chain Monte Carlo method. Multiple imputation is not sufficient to control for bias due to missing data, and MCMC is known to result in bias (N). 3.3: In general reasons for missingness were similar between groups, but the placebo group had disproportionately large missingness of data compared	4.1: Tools used for measurement are valid and appropriate tools to measure outcomes (N). 4.2: Same measurement methods and thresholds were used in all groups and measurements were done at similar time points (N). 4.3: Double-blind experiment (N). Overall: Low risk	5.1: Data produced and analyzed were consistent with the pre-specified analysis plan per the protocol (Y). 5.2: Primary and secondary outcome domains were measured in multiple ways, and data from all measures was appropriately included (N). 5.3: Eligible reported results for the outcome measurement correspond to all intended analyses (N). Overall: Low risk	High risk

			to treatment groups (PY). 3.4: Disproportionately high non-completion by participants in placebo group, with withdrawal by subject and other being most common reasons (PY) Overall: High risk			
Efficacy and Safety of Upadacitinib vs Dupilumab in Adults With Moderate-to-Severe Atopic Dermatitis	1.1.: Stratified randomization was noted; a randomization number used to encode each patient's placement based on a randomization schedule (PY). 1.2: Double-blind was mentioned, but blinding was not discussed further. Clinical trial registration indicated quadruple masking, but no	2.1: Although participants were blinded, those in the upadacitinib group disproportionally experienced acne; acne is a well-known side effect of upadacitinib (PY). 2.2: See 2.1 2.3: No indication that blinding was compromised. Rescue therapies were used as outlined in protocol (PN). 2.6: ITT analysis was conducted, and is	3.1: Population for ITT analysis was randomized. Greater than 5% and less than 20% of patients were discontinued from the study (N). 3.2: Non-responder imputation, multiple imputation, and per-protocol analyses were used as sensitivity analyses and are insufficient to correct for bias (PN).	4.1: Tools used for measurement are valid and appropriate tools to measure outcomes (N). 4.2: Same measurement methods and thresholds were used in all groups and measurements were done at similar time points (N). 4.3: Double-blind, double-dummy experiment (N). Overall: Low risk	5.1: Data produced and analyzed were consistent with the pre-specified analysis plan per the protocol (Y). 5.2: All outcome measurements and analyses were included in either the main paper or the supplementary appendix (N). 5.3: Eligible reported results for the outcome measurement correspond to all intended analyses (N).	Overall: Some concerns

	further information (PY) 1.3: No suspicious or excessive baseline imbalances or similarities noted (N) Overall: Low risk	considered appropriate (Y) Overall: Low risk	3.3: Substantial proportion of individuals withdrawing from the study or being lost to follow-up. No further information regarding their reasons was given. (PY) 3.4: Similar proportion of missing data per reason for missing data. It is unlikely that missingness of outcome depends on true value (N). Overall: Some concerns		Overall: Low risk	
Upadacitinib plus topical corticosteroids in atopic dermatitis: Week 52 AD Up study results	1.1: Stratified randomization occurred using an interactive response technology system (Y). 1.2: Double-blind and double-blind extension indicated, but	2.1: The study was double-blinded and quadruple masked. Differences between treatments would not be readily obvious. Placebo was used (N). 2.2: See 2.1. 2.6: ITT analysis was conducted, and is	3.1: Greater than 5% and less than 20% of patients were discontinued from the study (N). 3.2: Mixed-effect model with repeated measures was used, and was	4.1: Tools used for measurement are valid and appropriate tools to measure outcomes (N). 4.2: Same measurement methods and thresholds were used in all groups and measurements were	5.1: Data produced and analyzed were consistent with the pre-specified analysis plan per the protocol. Any *post-hoc* analyses were declared (Y). 5.2: Primary and secondary outcome domains were	Some concerns

	exact method used for blinding is unknown (PY). 1.3: No suspicious or excessive baseline imbalances or similarities noted (N) Overall: Low bias	considered appropriate (Y) Overall: Low risk	sufficient to account for bias due to missing information. Overall: Low risk	done at similar time points (N). 4.3: Double-blind experiment (N). Overall: Low risk	measured using multiple scales, and data from each measure was included (N). 5.3: No information regarding adjustment strategy indicated (NI). Overall: Some concerns	
Efficacy and Safety of Oral Janus Kinase 1 Inhibitor Abrocitinib for Patients With Atopic Dermatitis: A Phase 2 Randomized Clinical Trial	1.1: Randomization conducted using a computer-generated randomization schedule and assignment was done through an Interactive Voice Response System (Y). 1.2: Study record detail and protocol indicated that study was double-blinded and triple masked (Y). 1.3: No suspicious or excessive baseline imbalances or similarities noted (N).	2.1: Participants were blinded throughout the study and placebo was used (N). 2.2: It was not stated whether those delivering the intervention were masked (NI). 2.3: There is no evidence of unintended changes to treatment and use of rescue medications were predetermined in study protocol (PN). 2.6: Modified ITT analysis was conducted, and is considered appropriate (Y).	3.1: Population for ITT analysis was randomized. Greater than 20% of patients were discontinued from the study (N). 3.2: Modified ITT and sensitivity analyses were performed for missing data. Models used for handling missing data include generalized linear mixed model and mixed-effects model; these models did not use imputation (PY). Overall: Low risk	4.1: Tools used for measurement are valid and appropriate tools to measure outcomes (N). 4.2: Same measurement methods and thresholds were used in all groups and measurements were done at similar time points (N). 4.3: Double-blind experiment, outcome assessors were masked (N). Overall: Low risk	5.1: Data produced and analyzed were consistent with the pre-specified analysis plan per the protocol (Y). 5.2: All outcome measurements and analyses were included in either the main paper, study record detail, or the supplementary appendix (N). 5.3: Eligible reported results for the outcome measurement correspond to all intended analyses (N). Overall: Low risk	Low risk

	Overall: Low risk	Overall: Low risk				
Efficacy and Safety of Multiple Dupilumab Dose Regimens After Initial Successful Treatment in Patients With Atopic Dermatitis: A Randomized Clinical Trial	Randomization conducted using "predefined random number sequence with block size of 5 within each combination of stratification factors" (Y). 1.2: Study record detail indicated double-blinded and triple masked (including participant, investigator, and outcome assessor). The paper further indicated that all individuals involved were blinded except for the statistician who conducted the randomization; the statistician was not involved in the project in any other way. Study drug kits were blinded and coded	2.1: Study was double-blind and triple masked, including participant masking. Placebo was used (N). 2.2: Although the research protocol does not indicate that care providers were masked, the paper states that all involved individuals were blinded; additionally, it is indicated that the drug kits were blinded and interventions were replaced with identical placebos when intervention was not administered (PN). 2.6: ITT analysis was conducted, and is considered appropriate (Y) Overall: Low risk	3.1: Approximately 11% of participants were discontinued from the study (PN). 3.2: Multiple sensitivity analyses were conducted for placebo group and all intervention groups. Bias due to missing outcome data is unlikely (PY). Overall: Low risk	4.1: Tools used for measurement are valid and appropriate tools to measure outcomes (N). 4.2: Same measurement methods and thresholds were used in all groups and measurements were done at similar time points (N). 4.3: Double-blind experiment, outcome assessors were masked (N). Overall: Low risk	5.1: Measures and statistical analyses conducted in study were consistent with the pre-specified analysis plan (Y). 5.2: Multiple scales were used for both primary and secondary outcomes, and data was provided for all measures (N). 5.3: Eligible reported results for the outcome measurement correspond to all intended analyses (N). Overall: Low risk	Low risk

	with a numbering system (PY). 1.3: No suspicious or excessive baseline imbalances or similarities noted (N) Overall: Low risk					
Dupilumab does not affect correlates of vaccine-induced immunity: A randomized, placebo-controlled trial in adults with moderate-to-severe atopic dermatitis	1.1: Stratified randomization conducted using interactive response technology system (Y). 1.2: Study record detail indicated that study was double-blinded and triple masked (including participant, investigator, and outcome assessor) (Y). 1.3: No suspicious or excessive baseline imbalances or similarities noted (N) Overall: Low risk	2.1: Study was double-blind and triple masked, including participant masking. Placebo was used (N). 2.2: The study record detail did not specify whether care providers were masked (NI). 2.3: No indication was made as to whether there were deviations due to trial context (NI). 2.6: ITT analysis was not explicitly discussed; all individuals who were discontinued from the study were included in efficacy analyses (PY).	3.1: Population for ITT analysis was randomized. Greater than 5% and less than 20% of patients were discontinued from the study (N). 3.2: Last observer carried forward was used for missing information, and is not considered sufficient to account for bias due to missing information. 3.3: Inconsistencies between trial groups related to missing information may indicate that	4.1: Tools used for measurement are valid and appropriate tools to measure outcomes (N). 4.2: Same measurement methods and thresholds were used in all groups and measurements were done at similar time points (N). 4.3: Double-blind experiment, outcome assessors were masked (N). Overall: Low risk	5.1: Data produced and analyzed were consistent with the pre-specified analysis plan per the protocol (Y). 5.2: All outcome measurements and analyses were included in either the main paper, study record detail, or the supplementary appendix (N). 5.3: Eligible reported results for the outcome measurement correspond to all intended analyses (N). Overall: Low risk	High risk

		Overall: Some concerns	missingness is due to true value (PN). 3.4: Differences between groups are substantial, and likely indicate that missingness is due to the true value. Overall: High risk			
Abrocitinib versus Placebo or Dupilumab for Atopic Dermatitis	1.1: Randomization conducted using center-based permuted blocks (Y). 1.2: Study record detail indicated that study was double-blinded and quadruple masked (Y). 1.3: No suspicious or excessive baseline imbalances or similarities noted (N) Overall: Low risk	2.1: Study was double-blind and quadruple masked. Placebo was used (N). Although the individuals providing training to the carers were unblinded, the carers were blinded. It is unlikely that blinding was broken during training (PN). 2.6: A modified ITT analysis was conducted, and is considered appropriate (Y) Overall: Low risk	3.1: Data for outcomes was available for nearly all participants randomized. The number of individuals that discontinued from the study was less than 5% (Y). Overall: Low risk	4.1: Tools used for measurement are valid and appropriate tools to measure outcomes (N). 4.2: Same measurement methods and thresholds were used in all groups and measurements were done at similar time points (N). 4.2: Outcome assessors were masked (N). Overall: Low risk	5.1: Data produced, analyzed, and reported were consistent with the pre-specified analysis plan per the protocol (Y). 5.2: All outcome measurements were included for each pre-specified outcome measurement used (N). 5.3: Eligible reported results for the outcome measurement correspond to all intended analyses (N). Overall: Low risk	Low risk
Dupilumab Provides	n/a	n/a	n/a	n/a	n/a	High risk due to

Favorable Safety and Sustained Efficacy for up to 3 Years in an Open-Label Study of Adults with Moderate-to-Severe Atopic Dermatitis						open label study.
Efficacy and safety of abrocitinib versus dupilumab in adults with moderate-to-severe atopic dermatitis: a randomised, double-blind, multicentre phase 3 trial	1.1: Block randomization was conducted using an interactive response system (Y). 1.2: Study record detail and protocol indicated that study was double-blinded and quadruple masked (Y). 1.3: No suspicious or excessive baseline imbalances or similarities noted (N). Overall: Low risk	2.1 and 2.2: Participants were blinded throughout the study. However, the two treatment groups were administered differently (oral and subcutaneous). Although, an interactive response technology was used to dispense tamper-free packaging, it is still possible that carers could determine the treatment during follow-up appointments, particularly in cases where subcutaneous injections were	3.1: Population for full analysis set was randomized. Greater than 5% and less than 20% of patients were discontinued from the study (N). 3.2: Multiple imputation was used as a sensitivity analyses, and is considered insufficient to correct for bias due to missingness of data (PN). 3.3: Due to the stratification by disease severity, the similarity in	4.1: Tools used for measurement are valid and appropriate tools to measure outcomes (N). 4.2: Same measurement methods and thresholds were used in all groups and measurements were done at similar time points (N). 4.3: Outcome assessors were masked (N). Overall: Low risk	5.1: Data produced and analyzed were consistent with the pre-specified analysis plan per the protocol (Y). 5.2: All outcome measurements and analyses were included in either the main paper, study record detail, or the supplementary appendix (N). 5.3: Eligible reported results for the outcome measurement correspond to all intended analyses (N). Overall: Low risk	Some concerns

		delivered on-site rather than by the participant (PY). 2.3: Not enough information was provided to adequately assess this domain (NI). 2.6: A full analysis set was used and is considered appropriate (Y). Overall: Some concerns	baseline characteristics between treatment groups, and the similarity between groups regarding missingness of data, it is unlikely that missingness of data was due to the true value (PN) Overall: Low risk			
Abrocitinib versus Placebo or Dupilumab for Atopic Dermatitis	1.1: Randomization was conducted using an interactive response system and center-based randomly permuted blocks (Y). 1.2: Study record detail and protocol indicated that study was double-blinded and quadruple masked (Y). 1.3: No suspicious or excessive baseline	2.1: Participants were blinded throughout the study and placebo was used (N). 2.2: Both treatments and its matched placebo was blinded and administered similarly. Treatments and matched placebos were administered by an unblinded administrator who would not participate in any other study-related procedures; it is unlikely that this	3.1: Population for full analysis set was randomized. Greater than 5% and less than 20% of patients were discontinued from the study (N). 3.2: Multiple imputation was used as a sensitivity analyses, and is considered insufficient to correct for bias due to missingness of data (PN).	4.1: Tools used for measurement are valid and appropriate tools to measure outcomes (N). 4.2: Same measurement methods and thresholds were used in all groups and measurements were done at similar time points (N). 4.3: Outcome assessors were masked (N). Overall: Low risk	5.1: Data produced and analyzed were consistent with the pre-specified analysis plan per the protocol (Y). 5.2: All outcome measurements and analyses were included in either the main paper, study record detail, or the supplementary appendix (N). 5.3: Eligible reported results for the outcome measurement correspond to all	Some concerns

	imbalances or similarities noted (N). Overall: Low risk	would break the blinding of carers (PN). 2.6: A full analysis set was used and is considered appropriate (Y). Overall: Low risk	3.3: The participants were not stratified by disease severity. Not enough information to properly assess this domain (NI). 3.4: Due to the similarities between reasons for non-completion and unlikeliness that the circumstances of the trial contributed to the missingness of the data, it is unlikely that missingness depended on the true value (PN). Overall: Some concerns		intended analyses (N). Overall: Low risk	
Efficacy and Safety of Abrocitinib in Patients With Moderate-to-Severe Atopic Dermatitis A Randomized Clinical Trial	1.1: Randomization was conducted using an interactive response system (PY). 1.2: Study record detail and protocol indicated that study was double-	2.1: Participants were blinded throughout the study and placebo was used (N). 2.2: Individuals delivering the treatment were blinded and both the	3.1: Population for full analysis set was randomized. Greater than 5% and less than 20% of patients were discontinued from the study (N). 3.2: Although imputation was	4.1: Tools used for measurement are valid and appropriate tools to measure outcomes (N). 4.2: Same measurement methods and thresholds were used in all groups and	5.1: Data produced and analyzed were consistent with the pre-specified analysis plan per the protocol (Y). 5.2: All outcome measurements and analyses were included in either the	Low risk

	blinded and quadruple masked (Y). 1.3: No suspicious or excessive baseline imbalances or similarities noted (N). Overall: Low risk	treatment and study kits were blinded (N). 2.6: A full analysis set was used and is considered appropriate (Y). Overall: Low risk	used as one type of sensitivity analysis, linear mixed-effect models and tipping point analysis was also used, and is considered sufficient (PY). Overall: Low risk	measurements were done at similar time points (N). 4.3: Outcome assessors were masked (N). Overall: Low risk	main paper, study record detail, or the supplementary appendix (N). 5.3: Eligible reported results for the outcome measurement correspond to all intended analyses (N). Overall: Low risk	
Safety and Efficacy of Upadacitinib for Atopic Dermatitis in Japan: 2-Year Interim Results from the Phase 3 Rising Up Study	1.1: Randomization was mentioned but no further details were provided (NI). 1.2: Study record detail indicated that study was double-blinded and quadruple masked (PY). 1.3: No suspicious or excessive baseline imbalances or similarities noted (N). Overall: Low risk	2.1: Participants were blinded throughout the study. Placebo was used and participants were rerandomized into treatment groups (N). 2.2: Study record detail indicated that carer was masked (PN). 2.6: Not enough information was provided (NI). 2.7: Not enough information was provided (NI). Overall: High risk	3.1: Population was randomized. Greater than 5% and less than 20% of patients were discontinued from the study (N). 3.2: No sensitivity analyses were performed and were reported as is (N). 3.3: Not enough information (NI). 3.4: Not enough information (NI). Overall: High risk	4.1: Tools used for measurement are valid and appropriate tools to measure outcomes (N). 4.2: Same measurement methods and thresholds were used in all groups and measurements were done at similar time points (N). 4.3: Outcome assessors were masked (N). Overall: Low risk	5.1, 5.2, 5.3: Not enough information was provided to sufficiently assess this domain. Overall: High risk	High risk