Janus Kinase Inhibitors for Rheumatoid Arthritis: Effectiveness and Value
Response to Public Comments on Draft Evidence Report

November 26, 2019

Table of Contents

#	Comment	Response/Integration
Manufacturers		
AbbVie		
1.	ICER's model is limited to a single (homogenous) hypothetical cohort of conventional DMARD inadequate responders (csDMARD-IR), which does not reflect the larger, more complex real-world patient population for which payers provide pharmacy benefits. Only one of the five phase III upadacitinib trials is included in the economic modeling. Notably, the analysis excludes two relevant patient types, i.e., biologic-inadequate responders, and mono-therapy patients, which are believed to each make up more than 40% of targeted immunomodulator treated RA patients in practice.	Our model focuses on conventional DMARD inadequate responders, as this is the cohort in the majority of the key clinical trials. In addition to the base case, we assess JAK inhibitors in TIM-experienced patients. TIM monotherapy is not standard of care.
2.	The model assigns estimated values for HAQ-DI score improvement to generate its measure of effectiveness (QALYs). HAQ-DI is but one endpoint among equally important measures to assess the clinical benefits of treatment. Suppressing disease activity and protection against radiographically detectible joint damage are the other two major, independent, and equally important goals for the treatment of RA. Radiographic progression is excluded from the ICER model. Additionally, clinical trial outcomes important to patients such as pain, fatigue, onset of effect are omitted from the model.	We agree that there are many important potentially measurable endpoints to assess the clinical benefits of treatment. Radiographic outcomes, although important, are often not significantly changed at three and six months follow-up—the typical duration of the trials—and are therefore frequently not reported. In addition, clinical outcomes such as pain and fatigue are not uniformly reported in a way that allows for comparison across trials.
3.	In keeping with ISPOR recommended best research practices, we recommend the use of empirically generated data rather than estimates that rely on assumptions to calculate data such as EQ-5D results.	The utility scores used in the model were based on health state evaluations made by a United States general population sample using the EuroQol (EQ-5D) index. We did not have empirically generated utility data from the trials, or longer-term EQ-5D data from patients on these treatments.
4.	The model relies on a series of assumptions to estimate endpoints needed to calculate QALYs. This sequential use of assumed data has not been validated for accuracy and risks the stacking of error and uncertainty to estimate QALYs.	We used the DAS28 because we had clinical trial data on the proportions of patients within different categories of disease activity based on the DAS28 at three months for all treatments included in line one.
5.	ICER uses DAS-28 scores and applies a formula originally derived by Stevenson et al. to estimate HAQ-DI improvement in biologic treated patients. Stevenson used EULAR categories to calculate and assign HAQ. However, ICER pools 2 of the DAS-28 categories, and omits DAS-28 change scores that are a necessary component to calculate EULAR response. Testing these substantial adaptations is needed demonstrate validity.	Because we could not find a robust DAS28-to-HAQ mapping algorithm at three or six months for all treatment strategies included, we used a mapping algorithm from EULAR to HAQ. In the absence of patient data on change in the DAS28 from baseline, we mapped the DAS28 disease activity categories into EULAR response categories. We acknowledge that this mapping has not been validated, but we feel that it was the best use of available data.
6.	The ICER model uses a Markov cohort framework which in recent years has fallen out of favor by health technology assessment organizations because patient-level simulations	Markov cohort models are widely used in the field of health economics, including in evaluations by health technology assessment organizations.

#	Comment	Response/Integration
	or hybrid decision tree/patient-level simulations offer advantages over the older method. Patient-level simulation framework is recommended to attain more accurate results.	
7.	In sensitivity testing, we were able to confirm that the model using ICER's assumptions and its choices to estimate EQ-5D produce substantially different results from a model that uses empirically available data from the SELECT RA trials, and the most contemporary methods accepted by other health technology assessors. We also found that other inputs and structural characteristics that led to large variance in the output were cost of second-line treatment and collapsing MDA and HDA categories.	We acknowledge that models using different structures, assumptions, and inputs may produce substantially different results from the model reported on here. Without a detailed analysis of these differences, it is difficult for us to comment on the drivers of these differences or the appropriateness of alternative assumptions.
8.	The model uses an outdated method to estimate an older version of the EQ-5D. ICER has acknowledged that the Wailoo model used in the analysis competes with "more advanced" methods of calculating utility in RA. The older EQ-5D instrument is less sensitive, has shown to have a poorer fit, and produces lower estimates for quality adjusted life-years than the EQ-5D-5L. The EQ-5D-5L, preferably using primary sourced data, is recommended to improve model sensitivity and best fit.	As stated in the Draft Evidence Report: "Although the Wailoo et al. relationship produces a higher utility within the HAQ range of 1.0 to 1.5, the change in utility for this HAQ range was approximately 0.1, consistent with the change in the other model. Uncertainty in the Wailoo et al. mapping was evaluated in parameter sensitivity analyses." We are unaware of any analyses mapping the HAQ to the EQ-5D-5L.
Amgen		
1.	The evaluation of the disease activity score at three months is a key assumption and important factor for drug discontinuation. Clinical trial data evaluating disease activity continuously and as a response endpoint suggests drugs have a further 10% to 50% improvement from week 12 to week 24 (van Vollenhoven et al, 2012; Taylor et al, 2017; Fleischmann et al, 2019). Rheumatology real-world practice data suggests in moderate-to-severe RA patients on a biologic and conventional DMARD combination that disease activity measurements over 12 months occur infrequently and lead to a change in therapy in less than 50% of patients (Yun et al, 20189 Stever 2019). ICER needs to carefully consider the timing of the disease activity evaluation at three months, given the RWE variability and implications for early drug discontinuation in patients receiving effective treatment.	We agree that disease activity continues to decrease with continued therapy beyond three months, but guidelines and current practice—as described by experts—recommend switching at three months if remission or low disease activity is not achieved.
2.	The proposed mapping of DAS to utility is a complex three-step process with inconsistent assumptions based on intended use of the disease activity measure. In the original ICER RA model (2017), ACR score was mapped to EULAR score. This is a more credible approach of the applied mapping since both these scores incorporate both the change in disease activity and the current state. The proposed mapping of DAS28 to EULAR misses the opportunity to incorporate the dynamic nature of patient change by using a static value. The value of DAS28 is the	We are unaware of any direct mappings of DAS28 scores into a utility measure.

#	Comment	Response/Integration
	continuous nature of the score. Clinicians use their clinical evaluation of the patient to determine a need to change therapy. The proposed use of DAS28 and conversion to EULAR using breakpoints circumvents this judgement by potentially disregarding an effective therapy in improving patients on the cusp of a lower disease activity category. In addition, the mapping to utility from DAS28 to EULAR to HAQ, as stated in the analysis plan likely biases the results due to the differences between DAS 28 and EULAR. We encourage ICER to provide a more precise assessment of disease activity directly to utility, using the continuous data, instead of relying on a multistep process that negates physician judgement.	
3.	In modeling adverse events, ICER is proposing only to include serious infections. The data source for these parameters pre-dates JAKs and thus, does not include JAK data, the drugs of interest (Singh et al, 2011). ICER needs to include additional sources for this safety input to best reflect all drugs in the analysis. JAKs have a set of adverse events that not only include serious infections such as herpes zoster, but also vascular events and lipid elevation that could cause significant resource use, patient disutility, and even death over time. Additionally, the approved JAKs have a black box warning in their package inserts pertaining to the above-mentioned adverse events. A model focusing on these drugs that does not include these adverse effects could be considered flawed.	We agree that other adverse events, such as vascular events, may be important in the long term and may be reasonable for patients to consider given black box warnings. However, these event rates are very low in short-term studies and may come from studies using various doses, making them difficult to include in a model with a one-year time horizon. In addition, because these events are uncommon, we believe they would be unlikely to substantially impact the results of the model.
4.	As RA is a chronic disease, and patients respond differently to products over time, the base-case assessment of the cost-effectiveness should reflect the chronicity of RA and should be longer than 12 months. As the consequences of disease have non-reversible effects on the joints, the impact of these deteriorations should be considered over at least a 10 year, if not a lifetime, time horizon. The 12-month model is additionally limited in the side effects that may occur over time, especially the important cardiovascular effects that have not been included in the model. As a lifetime model is incorporated as a sensitivity analysis, and is a more common assessment time period, ICER should make this more relevant analysis the primary assessment of the model.	As stated in the report, we chose to model these treatments over a one-year time horizon due to the uncertainty surrounding the long-term impact of these drugs, the number of subsequent lines of TIMs, and if and when patients transition to palliative care. We did model these treatments using a lifetime time horizon as a scenario analysis. We also felt that a lifetime horizon produced results that lacked policy relevance, given our focus in this review to compare the outcomes of initial biologic agents, a focus that gets greatly diffused in a lifetime model.
5.	The change in response rate between first-line therapy and second-line therapy is widely accepted to be 10%. The current draft report assumed a difference in response of 16%. The rate difference is quoted from an assessment of Swedish patients, potentially early adopters on TNF is only from 1999-2006. In addition, it appears the publication only provides the rates after the switch, and not in comparison to a first line to second line switch as is the suggested use	We used the 16% reduction in treatment efficacy following first-line failure because it was based on prior published data from RA patients.

#	Comment	Response/Integration
	in the Draft report. We recommend switching back to the 10% change in efficacy for the second-line basket of drugs.	
Bristol-Myers Squibb		
1.	Please note that all three JAK inhibitors, including tofacitinib, have a black box warning for thrombosis, in addition to other safety warnings. Of concern is that these adverse events (AEs) have not been included as part of the Analytic Framework (Section 2.1, Figure 1.1). BMS recommends that ICER includes all safety warnings in the model for an accurate and comprehensive safety assessment.	Thank you. We agree that thrombosis is an important adverse event and have added it to the analytic framework in Section 1.2. See the response to the comment from Amgen about including thrombotic events in the model.
2.	The report states that adalimumab was chosen as a comparator due to its extensive use in clinical practice for line 1 treatment after failure by a conventional DMARD. Based on this rationale, etanercept could have also been included in line one as it is both indicated, and widely used in line one. Furthermore, although not chosen for this analysis, other agents are used and indicated for line-one treatment and this is not clearly stated in the report. BMS recommends that ICER acknowledge in its report that the model does not fully reflect actual clinical practice and guidelines as they have chosen a simplified and focused update on JAK inhibitors alone. ICER should state that all targeted immune modulators indicated for first line are appropriate treatment choices for providers and patients."	Line-one treatment options include several TIMs of which adalimumab is one example. We chose to use adalimumab as the comparator in our report because it was directly compared to JAK inhibitors in clinical trials. We have added text to the report clarifying this and noting that other TIMs are indicated for first-line treatment.
3.	BMS recommends to clarify in the report how switches were assessed since a switching event (or lack thereof) contributes to overall costs and quality-adjusted life years (QALYs).	The report states: "Treatment switching was based on disease activity as measured by the DAS28-CRP value (Table 4.1), with those in remission and with low disease activity remaining on the same treatment after the first three months, while those with moderate/high disease activity switch to a subsequent line of therapy at the end of the first three-month cycle" and "...treatment switching was assumed to be to a market basket of TIMs with efficacy averaged across the TIMs." We have added text clarifying the TIMs included in the market basket.
4.	In addition, being that safety is a critical determinant for treatment persistence and may differ between TIMs, BMS recommends that ICER consider safety events and discontinuations as reported in the trials.	We included adverse events related to serious infection, aligning with approaches used in prior economic evaluations of RA treatments. We assumed a uniform rate of serious infection due to concerns that published rates associated with specific TIMs are uncertain estimates. As stated in the report, we included estimates of later treatment discontinuation due to other reasons such as loss of efficacy, serious adverse events,

#	Comment	Response/Integration
		and patient preferences based on data on RA patients in the CORRONA registry.
5.	A cost-effective analysis (CEA) model structured around the DAS28-CRP endpoint is biased in favor of DMARDs that artificially reduce CRP levels. Because JAK and IL-6 inhibitors interfere with the IL-6 signaling pathway, they affect CRP levels, regardless of changes in RA disease severity, therefore potentially overestimating the efficacy of JAKs and IL-6 inhibitors. BMS recommends to anchor the model on endpoints estimated using DAS28-ESR as opposed to DAS28-CRP because of JAK inhibitors' sensitivity to DAS28-CRP due to their unique mechanism of action.	Thank you for your comment. We chose the DAS28-CRP in part because of its use as the primary outcome in recent clinical trials in RA, but we acknowledge issues with the DAS28-CRP as a primary outcome. We have reported the other outcomes in the clinical section including an NMA for the ACR20/50/70 outcomes, which were available for all trials.
6.	The model assumptions (Table 4.4, pages 46 to 48), lists only JAKs and TNFis as being in the market basket based on the availability of DSR28 data at three months. However, in the economic input assumptions on page 53, abatacept, rituximab and tocilizumab are included. Average net annual prices for all of the therapies are used to determine the net estimated price. BMS recommends ICER to clarify which treatments are included in the market basket and to be consistent between cost and efficacy analyses. While BMS understands the use of a market basket for the analyses, therapies included should be the same for both cost and efficacy assumptions.	Thank you for your comment. We have added additional clarification about the market basket to Section 4. We included in the market basket all treatments for which we had comparable data on cost or outcomes.
Eli Lilly and Company		
1.	This choice of the comparator presents an immediate challenge for the cost-effectiveness analysis due to lack of head-to-head clinical data vs. adalimumab for most JAKi therapies and lack of adalimumab clinical data in the TIM-experienced population. Therefore, we believe this comparator is inappropriate for a cost-effectiveness evaluation, especially in the TIM-experienced population, as the choice is inherently limited. Given most clinical trials in rheumatoid arthritis (RA) were conducted versus active comparator arm, i.e., placebo + MTX, ICER should conduct an indirect comparison of JAKi therapies vs. cDMARD as proposed in the MAP. Indirect comparison of JAKi therapies to each other is possible using clinical measures of treatment response (e.g., primary endpoint in clinical trials - ACR) at 12 and 24 weeks of treatment which were reported in the clinical benefit section of the Draft Evidence Report. Possible differences in patient characteristics between clinical trial cohorts should be matched adjusted for the indirect comparison.	Both upadacitinib (SELECT-COMPARE) and tofacitinib (ORAL-STANDARD) have been studied in head-to-head trials with adalimumab. Indirect comparisons were done for the ACR outcomes (see Table 3.3), but there were insufficient data to perform NMAs/indirect comparisons for the other outcomes.
2.	In the limitations section (page 64) ICER has indicated the following "we chose to model the second line market basket of TIMs over a one-year time horizon, which we believe to be a time period that patients will remain on	As stated in the report, we chose to model these treatments over a one-year time horizon due to the uncertainty surrounding the long-term impact of these drugs, the number of subsequent lines of

#	Comment	Response/Integration
	TIMs irrespective of multiple switches." This assumption is inappropriate for an economic model in RA, which should "include consideration of the connection between the prevention of radiographic progression and downstream economic consequences, and [therefore it is important] to employ lifetime models wherever possible because a long time period is necessary to determine the true cost-effectiveness of agents that modify radiographic progression of RA. In doing so, it is hoped that such [evaluation] will provide optimal information to facilitate important decisions on resource allocation." The one-year time horizon leads to misleading results of the model. RA is a chronic progressive disease with expected lifetime use of DMARD treatments. Evaluating one-year incremental cost-effectiveness ratios underestimates the impact of DMARDs use on a healthcare system as was evident from the sensitivity analysis which included lifetime horizon and was reported in the Appendix Table E10 (page 142 of the Draft Evidence Report). We suggest ICER chooses a longer than one-year time horizon in the base-case analysis that is more appropriate for evaluation of a chronic disease such as RA that requires lifetime treatment. The chosen time horizon should be specifically mentioned in the conclusions (page 65 of the Draft Evidence Report). In case of contradictory results from the base case vs. sensitivity analysis, discussion of the impact of the chosen time horizon vs. the lifetime scenario results should be added to the limitations (page 64 of the Draft Evidence Report).	TIMs, and if and when patients transition to palliative care. We did model these treatments using a lifetime time horizon as a scenario analysis. However, we felt that the lifetime horizon produced results that lacked policy relevance, given that it tended to magnify the effect of second and subsequent lines of therapy. We explored options for modeling follow-on care to address the concern, but shortening the time horizon appeared to be the most reasonable and policy relevant choice, especially because our focus in this review is to compare the outcomes of initial biologic agents, a focus that gets greatly diffused in a lifetime model.
3.	We would like to point out an inaccurate assessment by ICER of the study findings and limitations. First, the study found that access restrictions were common: one-third of those studied had access restrictions to at least one biologic or targeted synthetic DMARD treatment through step therapy or prior authorization or both. And of those with restrictions, nearly 70% of people with RA and 79% of people with PsA were enrolled in plans that required step therapy with or without prior authorization. (Section 2)	Thank you for your comment. We think that your observation that two-thirds of plans do not have access restrictions is remarkable. We have added it to the report.
4.	Among individuals with RA whose plans require step therapy to their RA treatment, medication adherence was 18% lower and odds of treatment effectiveness were 17% less compared to people with RA who did not have access restrictions. The impact of step therapy among people with PsA was even higher: medication adherence was 27% lower, and the likelihood of treatment effectiveness was 25% lower compared to people with PsA in plans without access restrictions.	This suggests possible selection bias if adherence is the mechanism for lower effectiveness. The direction of causality is not clear with this study design.

#	Comment	Response/Integration
5.	Study results show that people in plans with step therapy have additional healthcare resource use throughout the course of an individual's coverage period. For example, individuals with RA whose plans required step therapy were three times more likely to be admitted to the hospital due to an infection, and nearly twice as likely to visit the emergency room during the study compared to people with RA in plans without step therapy restrictions. In addition, those with access restrictions filled prescriptions for glucocorticoids and nonsteroidal anti-inflammatory drugs (NSAIDs) more often, which could be an indication of poorly managed disease.	We note that there were no differences in overall rates of inpatient admission.
6.	We do not believe that the reported step therapy effect size is small as ICER has pointed out in the Draft Evidence Report. The study was conducted using patient-level data and had an appropriate sample size to detect statistical differences between the study cohorts (the study included 3,993 people with RA and 1,713 people with PsA). The difference between patients in plans by restriction level should be expected as selection into these plans should not be based on patients' characteristics. These differences should not be over-adjusted by propensity score matching as the purpose of the study was to measure the impact of these differences on treatment outcomes. Among the baseline characteristics, only urban residence, proportion of patients with diagnosis of chronic respiratory condition, diabetes, hypertension, and osteoarthritis showed a statistical difference between the study groups. Regression adjustments conducted in the study were sufficient to account for these differences between cohorts.	We agree that the data are at the patient level. We intended to highlight that these are administrative data rather than data collected by speaking with and examining the patient.
7.	Baricitinib coverage by UnitedHealthcare – Employer & Individual (UHC - commercial) described on page 16 of the Draft Evidence Report is outdated. As of November 1st, 2019, UHC has prior authorization criteria for baricitinib in alignment with the FDA label, allowing access to RA patients diagnosed with moderately to severely active RA that have a history of failure, contraindication, or intolerance to at least one TNF antagonist therapy. Also, baricitinib will gain preferred brand tier drug status with UHC- commercial effective January 1st, 2020. We ask ICER to update the coverage information for baricitinib with UHC.	Thank you. We have updated Section 2 with this information.
8.	ICER incorrectly reports a lower significance level for baricitinib ACR20 and ACR70 response rates in Table 3.5, page 32 of the Draft Evidence Report. The report should have '*' next to both ACR20 and ACR70 response rates denoting $p<0.001$ as reported in Genovese et al. (2016), Table S4 of Supplementary Appendix. We ask ICER to make this correction.	Thank you. We have corrected the error.

#	Comment	Response/Integration
9.	Appendix Tables: D5, D6, D10, and D14, do not report baricitinib data from RA-BUILD and RA-BEACON clinical trials. Below we provide references where these data were reported. Please note that publications on clinical trials usually have Supplemental materials or Appendixes that are provided in a separate file along with the main publication. We ask ICER to include the complete clinical data for baricitinib in the report.	Thank you. We have corrected the omission.
Genentech		
1.	A one-year time horizon is not appropriate for evaluating the long-term cost-effectiveness of interventions for RA. Best practices for cost-effectiveness modeling recommend adopting a time horizon long enough to capture all health effects and costs. Although ICER used a one-year time horizon to mitigate the uncertainty surrounding the number of subsequent lines of therapy, this assumption does not reflect real-world treatment patterns, outcomes, and costs. A one-year time horizon fails to account for the chronic nature of RA in which treatment with TIMs can be lifelong even if patients achieve remission. Therefore, to assess the long-term cost-effectiveness of an intervention, a lifetime time horizon should be used to reflect the disease course of RA and treatment with multiple lines of therapy.	As stated in the report, we chose to model these treatments over a one-year time horizon due to the uncertainty surrounding the long-term impact of these drugs, the number of subsequent lines of TIMs, and if and when patients transition to palliative care. We did model these treatments using a lifetime time horizon as a scenario analysis. However, we felt that the lifetime horizon produced results that lacked policy relevance, given that it tended to magnify the effect of second and subsequent lines of therapy, and our focus in this review is to compare the outcomes of initial biologic agents, a focus that gets greatly diffused in a lifetime model.
2.	The DAS28-ESR should be used to assess disease activity in the cost-effectiveness model rather than the DAS28 calculated using C reactive protein (CRP). Using the DAS28-CRP rather than the DAS28-ESR may overestimate remission rates for targeted immune modulators that target specific inflammatory cytokines, impacting the validity of the cost-effectiveness model. While DAS28-CRP and DAS28-ESR may overestimate response compared to other measures, targeted therapies can have a differential impact on CRP levels and ESR. Agents inhibiting interleukin-6 (IL-6) and Janus kinase (JAK) signaling lead to a rapid reduction in CRP levels while not affecting ESR to a similar extent. Although treatment guidelines do not distinguish between the DAS28-ESR and the DAS28-CRP, the threshold values to determine disease activity (e.g. remission, low disease activity) correspond to the DAS28-ESR. Therefore, applying the DAS28-CRP to the recommended threshold values in this update may overestimate the efficacy of JAK inhibitors, potentially mischaracterizing their value.	Thank you for your comment. We chose the DAS28-CRP in part because of its use as the primary outcome in recent clinical trials in RA, but we acknowledge issues with the DAS28-CRP as a primary outcome. We have reported the other outcomes in the clinical section including an NMA for the ACR20/50/70 outcomes, which were available for all trials.
3.	ICER should use the best available clinical and economic information from all TIMs to inform the second line market basket, including data submitted as academic-in-confidence during the course of this update. For example, incorporating data from clinical trials of Actemra® (tocilizumab) in TIM-experienced populations (e.g.	Thank you for your comment. The choice of TIMs to include in the market basket average was based on the availability of DAS28 data at three months after initiating a TIM. We have added additional clarification about the market basket to Section 4.

#	Comment	Response/Integration
	RADIATE, ROSE, SUMMACTA, ACT-STAR), submitted to ICER on July 12, 2019, can better reflect real-world treatment options. By including this data ICER can also better align the TIMs selected to inform the efficacy parameters with those informing the cost parameters of the model.	
4.	The "Executive Summary" and "Report-at-a-Glance" should highlight the limitations of the cost-effectiveness model and explicitly state that the results of the JAK inhibitor update should not be directly compared to the results of the 2017 RA report. For this update, ICER made substantial changes to the scope of the report and the structure of the cost-effectiveness model. The extent of changes and the limitations of these changes should be sufficiently to documented in various sections of the report and related materials. Adopting this recommendation can facilitate more informed interpretations and discussions by health care decision makers as they evaluate TIMs for the treatment of RA.	Thank you for this suggestion. We have added text to the report explicitly noting that the results of the current analysis are not directly comparable to the prior report due to changes in the model.
Janssen		
1.	A longer time frame, beyond the one-year horizon used in the base case, would be more appropriate for capturing the long-term benefits of disease modifying drugs and the effect of switching seen over the course of long-term treatment.	As stated in the report, we chose to model these treatments over a one-year time horizon due to the uncertainty surrounding the long-term impact of these drugs, the number of subsequent lines of TIMs, and if and when patients transition to palliative care. We did model these treatments using a lifetime time horizon as a scenario analysis.
2.	The report states that UHC designates infliximab as the preferred product compared to infliximab-dyyb. However, this does not reflect changes to UHC coverage made in 2019. Effective October 1, 2019, infliximab (Remicade) and infliximab-dyyb (Inflectra) are co-preferred for UnitedHealthcare commercial plans. Effective June 1, 2019, UnitedHealthcare Community Plan requires use of infliximab-dyyb (Inflectra) and infliximab-abda (Renflexis) prior to use of infliximab (Remicade).	Thank you. We have updated Section 2 with this information.
3.	The report also describes coverage decisions from select regional payers (i.e., Kaiser Permanente, Health Net, and Medi-Cal). However, there is variation among regional payers, with many regional payers offering both Remicade and biosimilars, and some that prefer the biosimilar over Remicade. Also, within a payer, there are coverage changes year to year with different products getting preferred or equal status. Considering the variation in coverage among regional payers, we recommend that the ICER report focus on the national insurers based on the greater number of covered lives represented.	Thank you for your comment. As part of all CTAF reviews, ICER includes coverage information for California-based insurers in addition to national insurers.
4.	Section 3.4, Pages 37-39, Biosimilars for RA, subsection "Infliximab Biosimilar (Inflectra/CT-P13/Infliximab-dyyb): In	Thank you for the suggestion. We feel that it makes more sense to keep all of the clinical

#	Comment	Response/Integration
	order to clarify this section to the reader, Janssen suggests keeping the paragraphs regarding PLANETRA and interchangeability together, which would mean to reverse the order of the final two paragraphs.	information together before the summary paragraph.
5.	We suggest pointing out that this describes the use of infliximab-dyyb in naïve patients, which is being shown as supportive data to the PLANETRA randomized study. To complement this real-world evidence in naïve patients, we suggest also to present real-world data in the setting of switching. Janssen recommends referring to a meta-analysis study previously submitted during the Draft Scoping Document comment period: • In a meta-analysis of 62 real world studies of non-medical switching from an originator anti-tumor necrosis factor (TNF) agent to a biosimilar, the reported annualized discontinuation rate of the biosimilar was 21% and was consistent between rheumatology and inflammatory bowel disease. Among those who discontinued, the switchback rate to the originator biologic among all discontinuers was 62% across all therapeutic areas and 71% in rheumatology specifically. Across all nine studies with control arms, the analysis revealed a statistically significant 18% increase in discontinuations in patients who switched compared with those who did not.	Thank you for the suggestion and reference.
6.	Literature on conducting economic evaluations in health care advocate using appropriate time horizons that reflect the full period over which the costs and effects are captured, which is often over the patient's lifetime. Previous existing models in RA, including ICER's 2017 model, have used a lifetime time horizon.	As stated in the report, we chose to model these treatments over a one-year time horizon due to the uncertainty surrounding the long-term impact of these drugs, the number of subsequent lines of TIMs, and if and when patients transition to palliative care. We did model these treatments using a lifetime time horizon as a scenario analysis.
Merck		
1.	Inclusion of all relevant biosimilar products in the report is essential to present an intact picture of the biosimilar landscape, provide more complete information, and raise the awareness of all biosimilar options among patients, providers, and payers.	Thank you. The goal of this section is to set up a discussion about the use of biosimilars in general during the Policy Roundtable at the Public Meeting. As noted in the scope and in the report, the intent is not to exhaustively evaluate the clinical data for biosimilars.
2.	The interchangeability discussion regarding biosimilar is premature and should not be a focus of this report, because FDA has not given 'interchangeability' status to any product. We believe the focus of discuss should be on bio-similarities between biosimilars and originators.	Thank you. We respectfully disagree. We think that interchangeability is part of the discussion about the uptake of biosimilars in the United States.
Pfizer		
1.	We feel that ICER is using several different methods to assess the products in this review and the entire biosimilar piece is lost as it is a separate topic from the assessment of	Thank you. We chose DAS28 CRP in part because of its use as the primary outcome in recent clinical trials in RA, but we acknowledge the issues with

#	Comment	Response/Integration
	the JAK inhibitors. This leads to confusion for readers of the report and no easy way to summarize the results for the different therapies in a concise way as each has their own nuances. Pfizer has recommended from the beginning that the use of the DAS28-CRP at three months endpoint was not advisable as it would lead to complications with comparing the various therapies. While ICER represents this as a lack of data on the part of tofacitinib and baricitinib, in fact tofacitinib has the most robust and longest-term clinical data available of the JAK inhibitors. So, while tofacitinib lacks the one precise endpoint at one particular timepoint that ICER chose to use for their model, it has a wealth of other clinical efficacy data available to make indirect comparisons.	the DAS28-CRP as a primary outcome. We have reported the other outcomes in the clinical section including an NMA for the ACR20/50/70 outcomes, which were available for all trials.
2.	DAS28-ESR to DAS28-CRP conversion then taints both the comparison of tofacitinib to the cDMARD arm as well as the indirect comparison of tofacitinib to adalimumab as it relies on the tofacitinib to cDMARD comparison. The literature does not support simply using the relative proportion from clinical trials without a model controlling for covariates and ignoring variation in one arm of comparators. Finally, looking into the Appendix where the calculation is described, the method is still not fully transparent or laid out. If all three DAS28-ESR score categories were simply multiplied by 2 or 1.5 then you would end up with over 100% of patients. So, it is unclear if the remission category is the only category where 2x is used, the low disease activity is where the 1.5x is used and the high disease activity is simply the remaining patients or if ICER approached this differently. With rounding errors, it is difficult to tell. Recommendation: Remove this conversion and either select a different endpoint to compare products or restrict the conversation about tofacitinib to the clinical efficacy section.	As mentioned above, we acknowledge the issues with the DAS28-CRP as a primary outcome and have reported other outcomes in the clinical section, but we believe this was the best alternative for these comparisons. We multiplied by 2x for remission, 1.5x for low disease activity, with the moderate/high disease activity proportion derived as the remainder when remission and low disease activity were summed. We have clarified the description of the conversion calculation in the Appendix.
3.	The following are contradictive Pg 55: “We were unable to draw a comparison between tofacitinib and adalimumab in our analyses due to a lack of comparable efficacy data…However, we comment on the value of tofacitinib relative to adalimumab based on the relative cost effectiveness of these TIMs when compared to conventional DMARDs in their respective trials.” • Pg 63: “As stated earlier, we were unable to compare the cost effectiveness of tofacitinib versus adalimumab due to a lack of data. However, we compared the outcomes of the two TIMs relative to their respective conventional DMARD comparators. The different values noted in the tables below for the cDMARD comparator arms directly reflect outcomes observed in	Thank you for your comment. We have made several edits to clarify and avoid contradictory language.

#	Comment	Response/Integration
	the adalimumab and tofacitinib clinical trials, respectively." • Pg. 64: "Results from Tables 4.15 and 4.16 demonstrate that the use of adalimumab or tofacitinib compared to conventional DMARDs results in marginally more QALYs at one year, at a higher cost." • Pg 71: "Results from the indirect modeling comparison of tofacitinib to adalimumab suggest that for the marginal benefit tofacitinib offers, a price much higher than adalimumab may not be justified."	
4.	It is unclear why ICER changed the time horizon of the cost-effectiveness model to one year from a lifetime. This approach is not supported by ICER own value framework. Recommendation: Returning to a lifetime time horizon and performing a sensitivity analysis for the one-year time horizon.	As stated in the report, we chose to model these treatments over a one-year time horizon due to the uncertainty surrounding the long-term impact of these drugs, the number of subsequent lines of TIMs, and if and when patients transition to palliative care. We did model these treatments using a lifetime time horizon as a scenario analysis. However, we felt that the lifetime horizon produced results that lacked policy relevance, given that it tended to magnify the effect of second and subsequent lines of therapy, and our focus in this review is to compare the outcomes of initial biologic agents, a focus that gets greatly diffused in a lifetime model.
5.	Recommendation: Remove any voting questions addressed to the panel that relate to baricitinib or tofacitinib unless there is direct head to head data available as ICER provides no indirect treatment comparisons for these therapies aligned with the literature or their own value framework. If they retain questions where there is head to head data, the questions should be reworded to address the fact that the judgement of the panel is to be based on the clinical trial data alone and using different endpoints by product as the DAS28 conversion is flawed.	If there is insufficient evidence, then the panel will vote that the evidence is insufficient to differentiate between the interventions. Note that we did provide indirect evidence using the ACR20/50/70 as in the last report.
6.	Upon review, Pfizer identified several inaccuracies in the reporting of clinical data for tofacitinib and infliximab-dyyb. These are listed in Appendix A with their exact location for ease of correction. Recommendation: Fix transcription errors and inaccuracies in the report regarding tofacitinib and infliximab-dyyb.	Thank you for pointing out the transcription errors. We have corrected them in the report.
7.	Several of these changes (e.g. changing the model horizon to one year instead of a lifetime horizon) are directly contrary to ICER's value framework. Any deviation from the value framework should be strongly justified. Additionally, while ICER does state that draft evidence reports are just that, by releasing a press release they invite media comment on the draft reports which may turn out to be inaccurate and create a wrong impression about the relative value of the products. Press releases on the current	Our reports provide the rationale for modeling decisions, especially when these differ from our reference case. For example, we state in the report our reasons for choosing to model these treatments over a one-year time horizon rather than lifetime.

#	Comment	Response/Integration
	draft evidence report repeated conclusions from the report about tofacitinib that are incorrect given the confusion in how the clinical trial results are represented as cost-effectiveness model results within the report. Recommendation: Provide more justification on major changes in scope of reviews and in particular provide stronger justification for changes to reports that deviate from the ICER value framework. Additionally, any press releases on draft evidence reports should solely be a call for public comment with corrections released when major changes are made from the draft to the final evidence report.	
Sandoz		
1.	We applaud ICER for including a biosimilar in the comprehensive list of targeted immune modulators with FDA indications for RA. But, we respectfully request that ICER review the clinical evidence for an additional biosimilar, ErelziTM (etanercept-szzs). This clinical evidence includes EQUIRA and EGALITY. Etanercept is one of the two more frequently used biologics for patients suffering from rheumatoid arthritis. Additionally, ErelziTM was approved by the FDA over three years ago and the European Medicines Agency in June 2017, with patients in the EU benefitting from treatment.	Thank you. As noted earlier in the response to comments, we have stated throughout this process that the goal of the biosimilar section is to ground discussion during the Policy Roundtable at the Public Meeting and not to exhaustively review the literature on biosimilars for RA. The biosimilar chosen is solely used as an example for a more general discussion about the role of biosimilars in the United States health care system.
2.	We once again recommend that when ICER evaluates the clinical evidence of biosimilars it must be taken in context with the totality of the evidence. A recent review by Coory and Thorton that applied the GRADE evidence criteria to biosimilar trastuzumab found that the totality of the evidence would be categorized as high-quality evidence. However, if the randomized trials were evaluated in isolation from the other studies, it could be mistakenly rated as medium-to-low quality. ICER should utilize the same approach as these authors when evaluating the evidence from biosimilars.	Thank you for the citation. Please feel free to make this point during the Policy Roundtable discussion at the Public Meeting.
3.	In ICER's discussion of Biosimilars for RA on pages 37-38, we disagree that significant cost reductions have not been observed in the US. Savings have been realized by patients, health systems, integrated delivery networks and payers when switching to biosimilars.	Thank you. We have re-worded the sentence a bit, but we feel that the overall message is accurate. For example, see Lyman et al. N Engl J Med 2018; 378:2036-2044 and Frank et al. N Engl J Med 2018; 378:791-793.
4.	We appreciate the current challenges with doing detailed economic analysis for biosimilars. In the future, we strongly recommend that biosimilar(s) be included in the detailed economic analysis, along with the other interventions, because the primary value of biosimilars rests in providing increased savings and access to biologic treatments. Additionally, the introduction of biosimilar competition to the market impacts the relative pricing of competing biologics in a drug class. Biosimilars have the same efficacy	Thank you. In our assessment, detailed analyses are not necessary. If biosimilars are assumed to have the same benefits and harms, but cost less, then they are cost saving.

#	Comment	Response/Integration
	and safety as their reference biologic but they differ in cost. Therefore, the inputs for the biosimilar for an economic analysis would be the publicly available cost data and the efficacy and safety data of the reference biologic. Also, in the future, we recommend that the biosimilar evidence paradigm be included in ICER's Value Assessment Framework in order to help assess the clinical and economic value of interventions.	
Clinical Societies		
American College of Rheumatology		
1.	We strongly encourage ICER to point out where these pathways are not in line with FDA approval, ACR guidelines, and/or best practices and where step therapy criteria are more restrictive than the drugs' FDA indications. Specifically, this report mentions requiring failure of more than two tumor necrosis factor (TNF) antagonists before access to a janus kinase (JAK) inhibitor or failure of multiple JAK inhibitors prior to using a TNF antagonist. This prioritizes the rebate status of treatments over their clinical appropriateness, which may thereby increase overall costs by imposing upon providers an obligation to prescribe first a drug(s) they believe is less likely to be effective for an individual patient. These additional costs have a significant impact on our health care system and should be considered and potentially factored into a drug's QALY rating.	We emphasize throughout the report how the pathways being evaluated do or do not fit with ACR recommendations. We emphasize that by following the step guidelines, that costs may increase when a provider may be obligated to first prescribe a drug that he/she perceives to be potentially less effective.

2.	The ACR would like to see more emphasis placed on the role for biosimilars in the cost and value discussion. We are concerned that comparing cost effectiveness to adalimumab (Humira) alone does not provide sufficient context. Reviewing Medicare payment limits, which are based on average sales price (ASP) data, shows the extent to which biosimilars are helping to reduce cost for infliximab. As shown in the table below, the payment limit of the originator drug, infliximab (Remicade), has dropped 24 percent since early 2018. Biosimilars infliximab-dyyb (Inflectra) and infliximab-abda (Renflexis) have experienced a similar decrease. In comparison, prices of both certolizumab (Cimzia) and abatacept (Orencia) –biologic drugs without direct competition from a biosimilar product –continued to climb during this period, increasing by nearly four percent and eight percent respectively. The ACR shares ICER's desire to see maximal value for cost in the targeted immune modulator (TIM) space, and we believe ICER could help promote uptake of biosimilar use, where appropriate, if more detailed cost-effectiveness data in light of these newer price points, were part of this report.	Thank you. We hope that the added emphasis happens during the Policy Roundtable discussion at the Public Meeting. We encourage you to raise these issues during the Policy Roundtable. As noted above, we did not feel that detailed analyses were helpful as the message is a simple one: if biosimilars have the same benefits and harms, but cost less than the reference drug, then they are cost saving and should be preferred.
Patient Advocacy Organizations		
Arthritis Foundation		
1.	While we understand the limitations in clinical trial data, we remain concerned that ICER is only looking at moderate to severe patients, and those who have not had adequate response to conventional DMARDs. We appreciate the recognition that Patient Reported Outcomes are important in assessing clinical effectiveness, and we encourage ICER to work closely with patient organizations like the Arthritis Foundation that are collecting this data. Our Live Yes! INSIGHTS program is designed to collect data from arthritis patients based on the PROMIS 29 measure set. The selection of these measures is backed by a Nominal Group Technique process with patients that allowed patients living with arthritis to compare Quality of Life tools in their relevancy to a life lived with arthritis. To-date we have received over 20,000 survey responses and are finalizing the first reports on our initial findings.	The population we included in our analysis was constrained by the trial data available for the comparisons of interest. We look forward to seeing the reports on your initial survey findings.
2.	Overall, we remain concerned that the choice of therapy RA represents is not a singular event but rather an iterative "best choice" scenario over many decades of life. RA patients will cycle over many therapies for decades, and the ultimate outcomes represent the summed experience on many drugs. Such outcomes are not captured in two-year clinical trial data. Likewise, payer decisions are incentivized by knowledge that the average patient remains on a given policy for under three years. The reliance of ICER on such short-term data supports this	Part of the rationale to model these treatments over a one-year time horizon was due to the uncertainty surrounding the long-term impact of these drugs, the number of subsequent lines of TIMs, and if and when patients transition to palliative care. We felt that these results better target our focus in this review of comparing the outcomes of initial biologic agents. We point this out in our report and describe the uncertainty

	incentive, at the expense of consideration for patients' long-term health. While ICER has attempted to model treat-to-target approaches in the current model, the data referenced reveal the underlying limitations of the present understanding of clinical practice. We remain concerned that the analyses as presented do not adequately acknowledge the uncertainty in the current literature.	around these analyses as well as presenting sensitivity and scenario analyses.
3.	The target population includes only severe disease activity, aligning with available clinical trial populations from studies used in this review. As in our previous comments, we believe younger patients and those with early/mild RA should be included as they likely differ in the secondary prevention benefits they receive from long term treatment. ICER uses a hypothetical homogenous population, which may be consistent with clinical trial populations, but it is not consistent with the majority of patients with RA. In our scoping document comments we strongly recommended that ICER run scenarios for particular subgroups based on biomarker results, or other patient or disease characteristics, and that this would be informative to potential decision-makers for both understanding the variance in value across treatment strategies and for preventing the harm that could be caused by gross oversimplification of complex sets of outcomes.	As you point out, we aligned our target population with the available trial data. We felt that extrapolation from these trials to those with early/mild disease would introduce great uncertainty and questions about the validity of such analyses.
4.	It remains unclear as to how ICER will account for the frequent occurrence of comorbidities in this population.	We are not aware of data showing the differential impact of these treatments on other specific comorbidities.
5.	We remain concerned about the reliance on QALYs, which we and nearly all other public commentators this and other chronic conditions, have noted is an insufficient method to determine cost effectiveness, particularly among patients with chronic diseases.	ICER believes that the QALY is a highly useful and informative measure of patient outcomes with a broad context and long-standing application. Importantly, the QALY reflects patient preferences for health states in a consistent and evidence-based manner and the use of it rewards treatments both for improving life expectancy but also for improving quality of life.
6.	ICER costs are based on wholesale drug costs that may differ practice-to-practice and between government and private insurers. ICER should provide contextualization in their report to take into account that actual costs may dramatically over or underrepresent the costs actually incurred by provider.	Our report states that our net price estimates are based on United States-level data across all insurer types.
Partnership to Improve Patient Care		
1.	We urge ICER to take steps to more accurately capture the actual RA patient population by using a micro-simulation model to run a series of different patient scenarios and assess cost-effectiveness for a set of atypical patients. This is especially important due to the range of patient types with this disease. Specifically, the variation in both the severity of the symptoms and the amount of time patients have been managing the disease has implications for a	The population we included in our analysis was constrained by the trial data available for the comparisons of interest. We did not have robust data to allow modeling at a more granular level. We do perform multiple sensitivity and scenario analyses to reflect the potential variability in the base-case results over ranges of input values.

	treatment's impact on individuals. ICER could then present the results as a range and highlight the importance of individual decision-making and the key drivers of value across treatment options for different types of patients. Many stakeholders commented on the heterogeneity of RA patients and provided ICER with detailed suggestions for how to capture this in its model. For example, the Arthritis Foundation suggested ICER run "scenarios for particular subgroups based on biomarker results, or other patient or disease characteristics." As noted above, despite overwhelming evidence of the need to reflect patient specificity and heterogeneity across the syndrome, ICER finalized a model based on just one homogenous population: adults in the US with severely active RA with inadequate response to conventional DMARDs and naïve to TIM therapy."	
2.	Patients also consistently made the point that it is important to capture the long-term nature of RA, emphasizing that it is important to maintain a long-term perspective on treatment since patients' experience and treatments can change substantially over the course of the disease. Though ICER acknowledged this feedback, it used a model that works on a three-month cycle of effectiveness and assumed no sequencing. It simply assumed that discontinuation or treatment failure leads to palliative care with no chance of remission, and the base case model runs for just one year. This model loses all of the nuance that patients and advocates encouraged ICER to capture by looking at the long-term nature of the disease.	We agree that long-term outcomes are essential. Our model does capture sequencing through the market basket. If patients fail to achieve remission or low disease activity at three months, they are switched to a new therapy. This happens again every three months until remission or low disease activity is achieved. We do not model palliative care in this report. That is one of the many changes from the prior report. As stated in the report, we chose to model these treatments over a one-year time horizon due to the uncertainty surrounding the long-term impact of these drugs, the number of subsequent lines of TIMs, and if and when patients transition to palliative care. We did model these treatments using a lifetime time horizon as a scenario analysis.
3.	Patients and advocacy groups highlighted the extreme importance of incorporating disease-specific patient-reported outcomes (PROs). Global Healthy Living Foundation went so far as to offer a patient-reported outcomes registry of nearly 20,000 people with arthritis, which CreakyJoints created through funding from the Patient-Centered Outcomes Research Institute (PCORI). The registry, Arthritis Power, collects real world data and patient reported measures that can be combined with clinical and payer data to provide a picture of the real-world experience of RA patients. Instead of incorporating disease-specific PROs in their model, ICER used a Markov model built around transitions and health states designed as proxies of disease activity measures (specifically DAS28). Similarly, the outcomes of the model are expressed primarily in terms of disease response rates (ACR20, ACR50	Thank you for your input. We have included the results from an important recent publication using CreakyJoints data, though they do not directly inform the model. Note that the prior report used the ACR20/50/70 in the model, but the current report does not.

	ACR 70, and HAQ-DI). ICER's model directly contradicts what the patient advocates suggested would be the most appropriate way to evaluate the value of new therapies for RA in practice.	
4.	Building on the previous point about lack of inclusion of patient-specific PROs, it is important to remember the, often, significant differences between the disease-specific and generic PRO. The primary purpose of the disease-specific PRO is to maximize the sensitivity of the tool to the health-related quality of life of the specific patient and disease under investigation. By contrast, the primary purpose of the generic PRO is to compare across diseases – for which shared symptom relevance may be very low – and to fit into pre-configured domains for translation into the discriminatory QALY measure. Asking patients to answer questions that are irrelevant is likely to alienate respondents and increase the potential for missing or inaccurate responses. Second, they are likely to miss issues that are a specific feature of the disease under study. As a result, generic scales lack the responsiveness needed to measure change associated with effective treatment.	In our report, we used the HAQ-DI, an RA disease-specific PRO, to calculate the EQ-5D scores that were used in the cost-effectiveness analyses. We encourage the inclusion of relevant disease-specific and generic PROs in research trials.
5.	RA is a condition that has substantial, high-quality RWE. The network meta-analysis used to quantify absolute treatment effects in this study is still limited to trial data even though the vast majority of agents being evaluated have been in use for years and are captured in available RWE. There have been a number of studies built around the use of real-world evidence both in RA more generally and in the evaluation of JAK inhibitors as a class more specifically. There are even examples of where RWE has been incorporated into network meta-analyses for RA. Other studies confirm how important RWE is in the evaluation of treatment in RA patients. The populations studied by RA RCTs are often very different than those populations of RA patients in the real world. RCT populations tend to be younger, tend to have had the disease for less time and have had fewer alternative treatments than patients in the real world.	Thank you for your input. This is a challenging area and we certainly try to incorporate real-world data where appropriate. Since the updated report was spurred by the expected approval of upadacitinib, we lacked real-world data for that intervention. We certainly appreciate the reduction in uncertainty about long-term safety and discontinuation rates that comes with high-quality observational data as well as the comparative effectiveness for outcomes that require longer follow-up than used in the pivotal RCTs.
6.	The use of utility data derived from Health Assessment Questionnaire (HAQ) scores, which in themselves are derived from changes in disease response rate, indicates a greater risk of dilution of effects in over-translation across sets of outcomes. The more steps of translation there are, the more loss of variance in samples can lead to underestimation of the effects of treatments. The choice may have been understandable in a context of limited use of patient reported outcomes in RA treatment trials, but there is significant available data of this type.	Because we could not find a robust DAS28-to-HAQ mapping algorithm at three or six months for all treatment strategies included, we used a mapping algorithm from EULAR to HAQ. In the absence of patient data on change in DAS28 from baseline, we mapped DAS28 disease activity categories into EULAR response categories. We acknowledge that this mapping has not been validated, but we feel that it was the best use of available data. We also performed sensitivity analyses including baseline HAQ.

7.	Another concern is the use of the Wailoo et al (2008) algorithm for translating HAQ to utilities, and not those developed more recently in Hernandez et al (2012). Multiple publications have highlighted that the latter is more accurate and has been used in more recent models. Looking at figure 2 in Hernandez et al (2013), the EQ-5D slope in the naïve linear model is less steep than both observed data and the mixture model which implies that the conversion algorithm used in the ICER model is likely to underestimate the positive impacts of treatment.	The Wailoo algorithm was used in our previous review. We originally chose to use the Wailoo algorithm as the pain variable needed for the Hernandez model was not reported for all drugs in our original scope. We believe that Figure 2 in Hernandez et al. shows that the linear and mixture models produce fairly similar EQ-5D scores except at the extremes of poor and good health.
8.	An additional concern with the HAQ translation is how it is used to generate estimates of mortality probability. The ICER model concentrates on the relationship between levels of HAQ and mortality but there is evidence that levels of change in HAQ from baseline decreases the probability of mortality. Exclusion of this factor may underestimate the value of successful treatment in the ICER model.	Mortality has a negligible impact in our model, which uses a one-year time horizon.
Patients Rising Now		
1.	The overview information provided about biosimilars is brief (five paragraphs), and the coverage information about infliximab-dyyb compared to the reference biologic is superficial and fails to explore the deep, broad, important complexities of the biosimilar situation in the U.S., including patent issues (i.e., so-called "patent thicket"), reimbursement issues that are complicating market penetration of biosimilars in the U.S. (including the so-called "rebate trap"), differential incentives patients and clinicians may be facing because of the benefit structures of various health plans (including Medicare), and different cost-sharing requirements between medical and pharmacy benefits.	Thank you for the suggestions. The goal of this section is to serve as a springboard for discussion during the Policy Roundtable at the Public Meeting. It is not intended to be a comprehensive review of either the clinical evidence or policy issues surrounding biosimilars. That is beyond the scope of the review. We encourage you to make the points you highlight here during the policy Roundtable discussion of biosimilars.
2.	While the draft report is supposed to be about JAK inhibitors for RA, infliximab-dyyb is a different class of treatment with label indications for eight different conditions, only one of which is RA. Thus, information provided about infliximab-dyyb in the draft report is not substantively useful for clinicians, patients, or policy makers.	See the response to your prior comment.
3.	It would be appropriate for the draft report's "Coverage Policies" section to include information about how the treatments of interest are covered by various short-term/limited duration health plans, Association Health Plans, and Health care sharing ministry coverage options. While we realize that some of those options may not include coverage or reimbursement for any prescription medicines – and may have very high deductibles or annual out-of-pocket limits – we believe that including them in ICER's discussion of coverage policies (even if to just note that the treatments would not be covered by various	We were unable to locate specific data on short-term/limited duration health plans, as well as the other plans mentioned. However, we have added additional information to Section 1.4 that emphasizes the large burden that patients covered by some health plans may face.

	plans), would provide a more robust and complete picture for U.S. patients, clinicians and policy makers. Therefore, we would like ICER to discuss how they might include such information in their coverage section. If such information will not be included, we ask ICER to explain why it will not do so.	
4.	The draft report states that infliximab-dyyb "has not yet been approved as interchangeable," which implies that such approval is pending without citing any information. However, according to news reports, only one product (not infliximab) has conducted a switching study, and that study has not been submitted to the FDA.	Thank you.
5.	If ICER's intent was to show that – for the purposes of the draft report – infliximab-dyyb is equivalent to the reference biologic, then that could be simply stated in the main section of the draft report, and any supporting information that ICER deems necessary should be provided in an appendix; and why is the CTAF going to discuss biosimilars? And will this discussion be about infliximab-dyyb specifically, or a general discussion about biosimilars, which, as we've noted above, is a very complicated matter involving patent law, as well as regulations and various market activities by biopharma companies, private payers, and Medicare? We would recommend that ICER delete this information from its report on JAK inhibitors, and if ICER determines that biosimilars are an important issue on which the organization has insights to contribute to the ongoing policy discourse on this issue, then it should create a report specifically about biosimilars, such as it has done for indication specific pricing.	The goal is to engender a general discussion of biosimilars during the Policy Roundtable. We will not focus on infliximab-dyyb. We present the data to help ground the discussion.
6.	RA can cause significant problems in other organs including the cardiovascular system, lungs, and eyes. Therefore, although we recognize that the clinical studies and metrics related to treating RA focus on joint damage (as well as some markers of autoimmunity and inflammation), we urge ICER to expand their person-focused perspectives in the report with greater discussion of those non-osseous manifestations of RA.	Thank you. We have added additional information about other outcomes associated with RA, but there are limited data on these outcomes in the RCTs and observational studies of the JAK inhibitors. This assessment is not intended to be a review of RA nor a clinical practice guideline.
7.	Related to access, and patient's needs and perspectives, we are concerned that the Societal Perspective analysis only includes productivity costs. Such a narrow analytical framework might be appropriate if "society" is only concerned with economic output, but we believe that society – and its perspectives – has a much broader view that includes people's non-working lives. We believe this perspective of ICER is consistent with its overall reliance on the QALY as a basic analytical tool, which as we and others have noted, discounts such important societal perspectives.	Our analyses strive to include not only productivity impacts, but also a broad set of indirect and non-health care impacts, as listed in the impact inventory in Table E1 of the Appendix. However, we are often constrained by the lack of data on the impact of treatments on these broader issues.

8.	We would like ICER to specifically comment on why it chose not to include the following societal perspective parameters listed in Appendix E that are important to people and society: Patient Time Costs, Unpaid Caregiver Time Costs, Transportation Costs, Cost of Uncompensated Household Production, Cost of Social Services, Impact of Intervention on Educational Achievement.	As stated above, we were unable to find any data on the impact of these treatments on any of these items.
9.	Since ICER has written about indication specific pricing in the past, we wonder why the draft report does not explore that aspect of pricing, cost, and affordability. We would appreciate ICER discussing this issue.	Thank you for your comment. Indication-specific pricing is often a part of our Policy Roundtable discussion at the Public Meeting.
10.	The draft report uses the term "compliance" in one instance, but we believe that "adherence" is a better word choice, which is used elsewhere in the draft report, since it reflects the shared decision making and team approach people with complex chronic conditions should have with their clinicians, whereas "compliance" has much more paternalistic overtones.	Thank you. We have corrected our inappropriate word choice throughout the document.
11.	We believe there is a typo in this sentence that contains a triple negative: "Additionally, we believe it is unlikely that these patients will not switch to upadacitinib…." and suggest that perhaps the final phrase should be "unlikely that these patients will switch to upadacitinib.	Thank you for pointing out this mistake. We have revised this sentence in the report.
12.	ICER's intended audience appears to be public policy decision makers, we suggest that ICER's reports present the Societal Perspective first, and then present its "Base Case" scenario as a subset of the Societal Perspective analysis. We would appreciate hearing ICER's thoughts on presenting its analyses in that order.	ICER's Reference Case specifies that the base case use the health care sector perspective, as it is most relevant in the US setting. We will continue to present the societal perspective as an additional analysis in all of our reports.
Research Organizations		
Innovation Value Initiative		
1.	Models developed through IVI's OSVP, such as the IVI-RA model, provide a starting point for such analyses by allowing flexible individual patient-level simulation of treatment sequences, with flexibility to account for heterogeneity in patient characteristics and outcomes. Rather than falling back on the narrower approach exhibited in the current analysis, we encourage ICER to work with organizations having access to the required evidence and necessary expertise to provide more meaningful and relevant information for decision makers – and ultimately, increase value for both patients and insurers.	Thank you for your comment and suggestion.
2.	For the current report on JAKs in RA, ICER chose to use a different model structure from that used in its 2017 evaluation of RA therapies. Although ICER provides a rationale for this decision, it raises the question: what would the results of the cost-effectiveness analysis have been if their previous model would have been used?	We did not attempt to use the previous model for this report. We have added text to the report explicitly noting that the results of the current analysis are not directly comparable to the prior report due to differences between these models.

3.	In the current model, ICER adopted a six-month EULAR-to-HAQ mapping algorithm for the three-month DAS28 disease activity categories, assuming remission to reflect "Good" EULAR response, low disease activity to reflect "Moderate" EULAR response, and moderate disease activity and high disease activity to reflect EULAR response of "None." This raises a question about the impact of this assumption on model estimates.	We adopted this approach because of the lack of a robust DAS28-to-HAQ mapping algorithm at three or six months and the absence of data needed for direct EULAR calculation. We acknowledge that this mapping has not been validated, but we feel that it was a reasonable use of available data.
4.	The representation of the sensitivity analyses is not informative. Figure 4.3, Figure 4.4, Table E7, and Table E8 seem to reflect the sensitivity of the QALYs and costs with upadacitinib at 1-year follow-up rather than the incremental QALYs and costs with upadacitinib versus adalimumab. In addition, one would expect a tornado diagram or table for a measure of cost-effectiveness, such as the incremental net-monetary benefit at a certain willingness-to-pay.	We agree that these were not as informative as incremental results. We have replaced the prior tornado diagrams in the report with a new table showing one-way sensitivity analysis results for cost per QALY.
5.	There seems to be an error in Table 4.18. The difference in total costs between upadacitinib and adalimumab according to the table equals $124,000 - $97,900 = $26,100, which is not in line with an incremental cost per QALY of $92,000 when the difference in QALYs is 0.699 – 0.693 = 0.006.	Thank you for pointing out this error.
6.	It is unclear whether the reported response rates with the second line market basket of TIMs in Table 4.5 and Table 4.6 already includes the 16% reduction (i.e. the 0.84 multiplier) or not. In other words, is the 22%, 14%, and 64% used in the model directly or are these estimates adjusted first? In general, the calculation of the treatment. responses with the second line market basket is unclear. We would welcome detailed information about the source information and calculations of the estimates provided.	The 22%, 14%, and 64% reported in Tables 4.5 and 4.6 are the values prior to applying the 0.84 multiplier. We apply the 0.84 multiplier once in the model for 2+ lines. That is, we do not apply another 0.84 multiplier for movement from second to third line therapy. We have revised the footnotes to these tables to clarify this. Thank you for pointing out this ambiguity.