Tolebrutinib for Secondary Progressive Multiple Sclerosis
Response to Public Comments on Draft Evidence Report

May 29, 2025

Table of Contents

#	Comment	Response/Integration
Manufacturers		
Sanofi		
1.	**<u>Underrepresentation of the importance of slowing disability progression not due to relapses</u>** The traditional cost effectiveness model structures in multiple sclerosis were initially built with relapsing disease in mind because the available treatment options were all for relapsing-remitting forms of the disease. The cost-effective prices derived from these models were heavily driven by the costs and disutility coming from relapse activity. So, treatment options that had larger impacts on relapse suppression would achieve higher prices which was intuitive. However, these models always struggled to capture the full direct and indirect costs of disability progression. This became very clear once treatment options for primary progressive disease showed success in clinical trials. The traditional Markov cost effectiveness structure was adapted by removing relapses from these models but the direct and indirect costs associated with disability progression were not reassessed. A clear example of the consequences of these choices is ICER's review of ocrelizumab for RRMS and PPMS. Ocrelizumab was the first ever option for primary progressive multiple sclerosis yet ICER's recommended cost effective price for ocrelizumab for RRMS was four times higher than ICER's recommended cost-effective price for ocrelizumab for PPMS despite the many treatment options already available for RRMS. **<u>Recommendation</u>**: We recommend that ICER should at a minimum reassess how they are capturing direct and indirect costs of disability progression so options that are first in class are not penalized if they are targeting slowing or reversing disability progression rather than targeting relapse prevention. Alternatively, ICER and all bodies using cost-effectiveness need to assess if this adaptation of the traditional RRMS Markov model structure is the most appropriate way to model treatments for multiple sclerosis that are not solely focused on relapse prevention. The best model for multiple sclerosis treatments might be one that moves away from the traditional phenotype definitions and instead allows the model to appropriately weight and value the impact on relapses and the impact on disability progression in absence of relapses.	As noted in this comment, previous ICER reviews (and other MS cost-effectiveness analyses) treated relapses as discrete events and assigned additional costs to them separately from health state costs. Given that relapse-related costs are not included in our analysis, it is unclear why and how a cost-effectiveness analysis focused on SPMS should reassess health state costs in this context. We recognize the importance of a treatment being the first available treatment. Although this is not captured in the cost-effectiveness modelling, it is captured in the "benefits beyond health" section. The ICER value framework identifies these "benefits beyond health" as important elements of any overall judgment on long-term value for money, and all ICER reports have separate sections in which evidence and information pertaining to these elements are presented. These elements, including the lack of other treatment options in the space are part of the broader elements of value that the appraisal committee will consider. We acknowledge the limited availability of cost data specifically for patients with SPMS stratified by EDSS level and agree that future research should aim to generate more up-to-date and detailed cost estimates for this population. Given that therapies aiming to prevent disability progression may offset healthcare costs by delaying progression into more severe and cost-intensive EDSS levels, it was essential to select cost inputs that appropriately reflect this dynamic. We opted to use the cost inputs from the ICER 2023 report, adjusted to 2024 USD.

		This source reports higher cost estimates than previously recommended, which we believe more accurately reflect the current standard of care. Additionally, the ICER-based estimates exhibit greater cost differentiation across EDSS levels, particularly as patients progress to more severe states. This better captures the potential for cost offsets associated with therapies slowing progression.
2.	**Evidence rating for Tolebrutinib** We appreciate ICER's partnership as we continue to provide data from the clinical trial programs once they are fully validated and are ready for dissemination. The GEMINI and HERCULES programs are robust, large programs and analysis was ongoing for several months after the phase 3 readout in September 2024. The data from GEMINI 1 and 2 and HERCULES were published in The New England Journal of Medicine just days before the release of the draft evidence report so we appreciate that ICER did not have full access to the data prior to the report's release. However, now that more data is available, we believe the P/I assessment rating from ICER does not reflect the true strength of evidence. ICER cited concerns about the lack of improvement in brain atrophy, limb mobility via the 9 hole peg test (9-HPT), and liver monitoring as reasons for their assessment for tolebrutinib. GEMINI 1 did demonstrate a nominally significant improvement in brain volume loss (BVL). While GEMINI 2 and HERCULES did not demonstrate statistically significant improvements in BVL, the BVL observed in the tolebrutinib arms in GEMINI 1 and 2 and HERCULES (about -0.7% from month 6 to end of study) was approaching what is observed in healthy people. It is important to note that teriflunomide, the active comparator in GEMINI has been shown to have significant benefit in BVL measurements versus placebo. While the importance of BVL in RRMS is fairly well understood, the importance of BVL in SPMS is not as well understood. Within HERCULES, liver enzyme elevations (>3xULN) were observed in 4.1% of participants receiving tolebrutinib compared with 1.6% in the placebo group, a side effect also reported with other BTK inhibitors in MS. A small (0.5%) proportion of participants in the tolebrutinib group experienced peak ALT increases of >20xULN, all occurring within the first 90 days of treatment. All but one case of liver enzyme elevations resolved without further medical	We appreciate the additional data provided about the effects of tolebrutinib on brain volume loss (BVL). However, although BVL is a marker for progression independent of relapses, particularly for RRMS, literature suggests that BVL is not as clearly associated with clinical outcomes in SPMS (Koch et al. Multiple Sclerosis Journal 2022, Vol. 28(4) 561 – 572), so it is not clear how the effects of tolebrutinib on BVL in the RRMS population (the target population in GEMINI 1 and 2) may translate to clinical outcomes in the SPMS population. While there is evidence that tolebrutinib slows progression of SPMS, there is also the potential of serious harm from liver toxicity. Although the harm may be mitigated by increased monitoring, clinical experts have indicated that the intense monitoring schedule in the clinical trial presents a large burden for patients and thus it may be difficult to fully implement into practice. Additionally, studies have shown that postmarket safety events, including black box warnings and drug withdrawals are common (Downing et al. JAMA 2017;317(18): 1854-1863; Lasser et al JAMA 2002;287(17):2215-20). Our understanding of the HERCULES trial is that the secondary outcomes were evaluated in a hierarchical design. Because the 9HPT outcome was not statistically significant, per the published protocol, outcomes evaluated after 9HPT such as 25FWT, CDI, and BVL should not tested for statistical significance to avoid Type 1

	intervention. Prior to the implementation of the revised study protocol with more stringent monitoring, one participant in the tolebrutinib arm experienced hepatotoxicity leading to a liver transplant and died due to post-operative complications. Once we identified a signal for hepatotoxicity in mid-2022, the stringency of monitoring was increased such that weekly monitoring was implemented between weeks 2 and 12, with monthly monitoring from month 3 until month 12 and quarterly monitoring through end of study. For limb mobility, the upper limb mobility as captured by 9-HPT was not statistically significant. However, lower limb function measured by the 25FWT was nominally significant (statistically significant but lower in the hierarchy than the failed 9-HPT endpoint). Across the full trial population (both treatment arms), only 25% of participants experienced ≥20% worsening in 9-HPT scores. **Recommendation**: With the additional data provided separately from these public comments given embargoes and space constraints, we recommend that ICER reassess the evidence rating and the impact of the liver monitoring schedule both on the evidence rating and model.	error. Thus, while there may have been numerical differences between the tolebrutinib and placebo groups, whether these represent statistically and clinically significant findings cannot be ascertained from the trial data. Finally, we note that the effect of tolebrutinib on patient-important outcomes such as health-related quality of life and cognition have not yet been reported or published. ICER's evidence rating reflects the overall net health benefit for the intervention under consideration, based on the certainty of evidence and the size of the net health benefit. A rating of P/I indicates that there is both the possibility of small to substantial net health benefit and the potential of net harm. Our rating acknowledges both the potential benefits of tolebrutinib (including substantial net health benefit), the uncertainties about the data, and the potential harm from hepatotoxicity.
3.	**Costs and other discrepancies in the model, report, and SAP** There appears to be a discrepancy in the costs used between the report, SAP, and model. The SAP referenced Hernandez (2016) interpolation of Kobelt data inflated using US BLS (2023a), however the report mentions the use of Hernandez (2016) interpolation of ICER 2020 Review which are much higher costs than the one mentioned in the SAP. Given the issues mentioned above around the model underrepresenting the costs of disability progression, please ensure the best data is being used and the SAP and report are consistent. There are also small discrepancies between the numbers seen in the model itself versus reported in the report that we cannot trace the source. **Recommendation**: Please relook at the costs used in the model in comparison to those in the report and SAP and ensure the correct costs are being used. Please also	As noted in the comment above, for the Draft Evidence Report, we updated the cost inputs proposed in the MAP to reflect those that more accurately represent the current standard of care. Specifically, we used cost estimates from Kobelt et al. (2006) and Bebo et al. (2022), as reported in the ICER 2023 review and inflated to 2024 USD. The draft evidence report previously referenced the inputs described in the MAP, which was incorrect; this has been corrected in the Revised Evidence Report. The cost inputs used in the model are consistent with the values reported in the Key Inputs table.

	confirm if the discrepancies between the output tables and the model have been resolved.	
4.	**Inclusion of CDI in the model** The draft report mentions a desire to include the nominal CDI results in the model structure. We support capturing disease improvements observed in both treatment arms. The London Ontario dataset, while historically valuable, unfortunately does not adequately reflect potential improvements. To address the need for more information about disability improvement, we derived annual transition probabilities using the HERCULES data that can be used as an alternative to individual arm EDSS progression source, in the cost-effectiveness model. However, this does have a significant limitation (aside from the classical trial duration constraint and restricted patient population limitations). The absence of EDSS 9 observations in HERCULES prevented us from constructing complete transition matrices. In the interest of transparency, Sanofi has provided the transition probabilities for EDSS 1-8 derived from HERCULES. These could be supplemented with London Ontario data for transitions not captured in our trial data. The advantage of this approach is that it would allow for straightforward modification of the US ICER model, by applying CDP and CDI Hazard ratios in a consistent way, to the relevant elements of the placebo matrix, thereby comprehensively reflecting the treatment's dual benefits: slowing disease progression while effectively capturing additional improvements. **Recommendation:** We recommend using the HERCULES placebo derived transition probabilities supplemented with the London Ontario dataset to fill in the EDSS 9 gap and applying the CDP and CDI hazard ratios for tolebrutinib to capture both the impact of tolebrutinib on disability progression slowing and disability improvement.	We reconstructed the model trace using the placebo arm's transition probability matrix provided by Sanofi as academic in confidence (AIC) data. These transitions now include disability progression and improvement. However, since the matrix did not include probabilities for transitioning to or from EDSS 9, we included EDSS state 9 without transition probabilities. Estimating these transition probabilities based on the London Ontario dataset would have required several assumptions, which would have introduced unnecessary uncertainty into the analysis. Future iterations of the model may incorporate movement to and from EDSS state 9 as data become available. To estimate treatment effectiveness, we applied hazard ratios for CDP. As stated in the draft report, our base-case model does not incorporate 6-month confirmed disability improvement (CDI), and this assumption remains unchanged. The 6-month CDI endpoint does not provide sufficient evidence to determine whether observed improvements represent long-term disease reversal or simply a temporary slowing of progression. Longer-term follow-up data would be necessary to confirm sustained improvement. CDI hazard ratios were included in a scenario analysis.
5.	**Major drivers of ICERs from the Cost Effectiveness Model** There are three further issues with the model structure that have a large impact on the resulting ICERs for tolebrutinib that should be revisited. The general mortality used is from 2019 while a 2023 version is available. The medical "unrelated" costs are unclear and no source is provided	More recent life table estimates are subject to mortality during COVID-19, which is the reason we used the life table from 2019.: On unrelated healthcare costs, ICER's standard approach is to include future healthcare costs in its assessments. Future

	disclosing where these are from. Finally, there seems not to be any half cycle corrections in the model. **<u>Recommendation:</u>** We recommend updating the mortality to the latest version, providing more clarity on the source of the medical "unrelated" costs and perhaps reassessing them in light of the underrepresented costs mentioned above, and applying half cycle corrections in the model.	health care costs include future costs related and unrelated to the condition under review, both of which represent real costs to the health system. Additional details can be found in ICER's *Reference Case for Economic Evaluations: Elements and Rationale*. https://icer.org/wp-content/uploads/2024/02/Reference-Case-4.3.25.pdf We have also added a citation in the revised evidence report to the source used for estimating unrelated healthcare costs (Jiao et al., 2021). Half-cycle correction: We appreciate the suggestion to apply a half-cycle correction. However, half cycle corrections have negligible impact on findings over a lifetime horizon. In this context, the added complexity does not meaningfully improve accuracy, and thus we opted not to apply the correction.