Tirzepatide for Type 2 Diabetes
Response to Public Comments on Draft Evidence Report

January 6, 2022

Table of Contents

#	Comment	ICER Response
Manufacturers		
Boehringer Ingelheim		
1.	Concerns related to the comparative clinical effectiveness with empagliflozin BI respectfully disagrees with the comparative clinical effectiveness rating of "C++" based on the assessment of net health benefit of tirzepatide compared to empagliflozin. The net health benefit assessment is based on extremely limited indirect comparison data and does not take into consideration well-established clinical outcomes that are relevant to T2DM treatments, thereby resulting in low certainty for the findings. BI recommends that the comparative clinical effectiveness rating should be "I" (insufficient), which is consistent with ICER's definition ("any situation in which certainty in the evidence is low"). BI provides the following reasons in support of the recommendation: A. Wide confidence intervals of the NMA estimates B. Limitations of the biomarkers used in representing the full range of T2DM treatment benefits C. Exclusion of cardiorenal metabolic benefits in NMA underestimates the value of empagliflozin	We appreciate the feedback about our comparative clinical effectiveness rating for tirzepatide compared with empagliflozin. Our rating of C++ is meant to reflect the fact that tirzepatide showed significant improvements in intermediate outcomes such as HbA1c, weight, LDL and SBP, but that because of the limited head-to-head comparison and lack of definitive cardiovascular and renal outcome data, there is greater uncertainty about whether tirzepatide has comparative, incremental, or substantial benefit compared with empagliflozin. Additionally, there is observational evidence that control of risk factors such as HbA1c, LDL and SBP is associated with improvement in cardiovascular outcomes (e.g., Rawshani et al, *N Engl J Med* 2018; 379:633-644; Colyaco et al., *Diabetes Care* 2011;34(1):77–83), and we have also added cardiovascular safety data from SURPASS-4, which shows evidence of tirzepatide's cardiovascular safety and a trend towards cardiovascular benefit. Furthermore, our modeling work from the draft report demonstrates that even assuming no direct cardiovascular benefit from tirzepatide, the QALY gains were still higher than empagliflozin, so it is reasonable to assume that tirzepatide is at least comparable to empagliflozin. We have clarified the reasons for our ratings in the revised evidence report.

2.	A. Limited indirect comparative data increases the uncertainty of NMA based treatment effects There are no head-to-head trials comparing tirzepatide and empagliflozin. For the assessment of net health benefit of tirzepatide versus empagliflozin, ICER developed quantitative, indirect comparisons using a Bayesian NMA for outcomes of change in HbA1c, weight, LDL, and SBP at 40 weeks in adults with T2DM. Estimating the relative treatment effects on HbA1c, weight, LDL, and SBP without head-to-head evidence impacts the precision of the estimates and increases the uncertainty of the comparative evidence. While the NMA leveraged available data, only 410 patients who received empagliflozin 25 mg (PIONEER 2, see Table 3.2 in the Draft Evidence Report), were included in the analysis. This is a significant underrepresentation of the population in the evidence base for empagliflozin, as this is approximately 2% (over 12,000 subjects in trial settings) of the overall empagliflozin population and does not take into consideration treatment with 10 mg empagliflozin. ICER acknowledges concerns with the scarcity of data and the resulting uncertainty surrounding the estimates/results in its Draft Evidence Report, for example on pages ES2, ES3, 18, 19, and 33. Describing the NMA, ICER states that "we have only moderate certainty about the results from the indirect comparison through the NMA, as tirzepatide and empagliflozin are compared through trials of three other drugs."(page 19) BI urges ICER to also emphasize that until additional and longer-term data is available, any assessment will not accurately capture the comparative value of tirzepatide and empagliflozin. The conclusion should therefore reflect these critical limitations with a low certainty in the evidence and result in a rating of "I", in line with ICER's own definition of the ratings.	We appreciate that our NMA comparing tirzepatide and empagliflozin was limited due to the data available linking the two drugs. We believe that we have accounted for that uncertainty in our evidence ratings. A rating of "I" is used in situations where we do not have sufficient evidence to do comparisons. In this case, although there are limitations to the NMA, we do have some ability to compare the two drugs. A rating of "P/I" is used when there is not enough evidence to make conclusions about a drug's effectiveness – that is not the case for tirzepatide. Furthermore, when one manufacturer suggests higher evidence ratings and another suggests lower ratings, it indicates that our chosen ratings likely appropriately reflect the current evidence.

3.	B. HbA1C and body weight alone do not capture key treatment benefits in T2DM ICER's evaluation assesses T2DM treatments based on glucose-lowering and weight modification therapies. These traditional biomarkers for health in the T2DM population do not correlate with the overall benefit demonstrated in studies of the SGLT2 inhibitor class, such as empagliflozin. Evaluating empagliflozin solely on its merits of a glucose lowering T2DM agent without accounting for its established CV benefits, underestimates the value of empagliflozin, undermining the integrity of the review. Empagliflozin has demonstrated efficacy and safety in clinical trials for the treatment of T2DM via glucose lowering and weight loss. However, the overarching value of empagliflozin extends beyond these intermediate measures of clinical outcomes. Modeling the relative value of empagliflozin based on a narrow set of biomarkers such as HbA1c and body weight does not provide assurance that its well-established clinical benefits are accurately reflected, especially given availability of long-term data. To conduct a fair and comprehensive comparative clinical assessment of T2DM treatments including empagliflozin, one should take into consideration each therapeutic agent's complete, proven vector of benefits. Despite diabetes being characterized by hyperglycemia, there are many dysmetabolic factors that lead to the multitude of comorbidities associated with T2DM. Among the most notable is CV disease, which is particularly diffuse in the T2DM population. This particular comorbidity is thought to relate to lipid metabolism which often precedes hyperglycemia by 5-10 years. The normalization of glucose levels in patients with T2DM and CVD has not successfully demonstrated a benefit in reversing or reducing CV events. In particular, two major T2DM trials, ACCORD and ADVANCE, failed to demonstrate that lowering HbA1c and blood glucose would reduce mortality. Changes in HbA1c and body weight do not adequately demonstrate an overall benefit to a multi-morbid T2DM population with regards to overall mortality, and major comorbid outcomes such as CV events, renal decline, and heart failure (HF), that determine survival and quality of life in the diabetes population. Conducting a comparative clinical assessment within this narrow view, as approached by the ICER evaluation, does	We appreciate this comment highlighting the additional benefits of SGLT2 inhibitor drugs beyond glucose and weight outcomes, including the cardiovascular and renal outcomes, and have qualitatively discussed these benefits for empagliflozin throughout the clinical effectiveness section of the report and supplement. These assessments have been factored into our evidence rating, as explained above.

	not accurately portray the complete clinical value of T2DM treatments, and in particular does not allow for a robust comparison with empagliflozin, given its established benefits in comorbid conditions of T2DM. BI emphasizes that given these limitations, certainty in the assessment for comparative clinical effectiveness of tirzepatide and empagliflozin is low, and therefore the rating should be "I".	
4.	C. CV and renal benefits are not adequately represented in ICER's evidence assessment CVD and chronic kidney disease (CKD) are common comorbid conditions with T2DM. Evaluating the effect of T2DM treatment must consider the impact on comorbid conditions, such as CVD and renal disease. The T2DM population is at a 2-5-fold increased risk of experiencing HF and approximately 45% of all HF patients have underlying T2DM. The risk of morbidity and mortality for T2DM patients increases with the presence of CVD and is compounded with the presence of renal disease. A systematic review of 57 global studies, covering more than 4.5 million T2DM individuals, documented that CVD had an overall prevalence of 32.2% and accounted for 50.3% of all deaths in this population. Additionally, an estimated 70% of healthcare costs in T2DM population is driven from macrovascular disease. A study in the NHANES adult T2DM population from 1999 to 2012 documented that the overall prevalence of CKD was 43.5% (95% CI, 41.6%-45.4%) based on estimated glomerular filtration rate (eGFR). Empagliflozin, an SGLT2 inhibitor, has demonstrated efficacy in CV and renal outcomes. Empagliflozin is indicated to reduce the risk of CV death plus hospitalization for heart failure (HHF) in adults with HF and reduced ejection fraction (HFrEF); to reduce the risk of CV death in adults with T2DM and established CV disease; and as an adjunct to diet and exercise to improve glycemic control in adults with T2DM. BI has submitted an application to FDA seeking a new indication based on the HFpEF data and, in September 2021, was granted FDA breakthrough therapy designation for HFpEF. Additional research is underway to assess its impact on both chronic kidney disease (CKD) and kidney function decline. EMPA-REG OUTCOME offers data on outcomes for T2DM comorbid conditions such as CVD and renal complications, HHF and total hospitalizations, for up to 5	See above answer.

	years of exposure to empagliflozin. The study, which examined the effect of empagliflozin as a treatment for T2DM patients at high risk for CV events receiving standard care, provides scientifically robust data for a mean of 3.1 years and over 780 outcomes (or events). In EMPA-REG OUTCOME, empagliflozin demonstrated a significant (14%) reduction in 3-point major adverse CV events (MACE), a 38% reduction in CV death, a reduction in the decline of glomerular filtration rate by 1.5ml/min/1.073m^2/year, a 35% reduction in HHF, and a 39% reduction in renal end points. This was achieved in a study designed to maintain glucose equipoise, which in the end demonstrated less than a 0.5% reduction in HbA1c over 3.1 years and a modest blood pressure reduction of approximately 3 mmHg SBP, while maintaining no change in heart rate, unlike the GLP-1RAs, which have been shown to increase heart rate in clinical trials. GLP-1RAs (such as liraglutide, and to a lesser extent semaglutide) have demonstrated CV benefit, but only have a minor impact on renal benefit and no effect on HF, despite greater glucose lowering and weight loss. Additionally, in EMPA-REG OUTCOME, time to CV benefit (a decrease in CVD and HHF) was observed within weeks of treatment initiation of empagliflozin, as compared to 12 months for GLP-1RAs, suggesting not just broader CV benefits for SGLT2 inhibitors, but also faster occurrence. Tirzepatide, a dual GIP and GLP-1RA, has only demonstrated efficacy as an antihyperglycemic agent. Regardless of data limitations for tirzepatide, the multitude of well-established and documented benefits of empagliflozin should be recognized and taken into consideration for a robust comparative analysis. BI recognizes that there is an ongoing CVOT for tirzepatide, which will provide robust data for a future comparison. However, for the current analysis, the Draft Evidence Report concludes that, “the cardiovascular outcomes trial for tirzepatide is ongoing and less mature; however, a meta-analysis of cardiovascular events for safety across the SURPASS trials showed no increase in cardiovascular events and a trend towards cardiovascular benefit.” ICER also repeatedly notes low levels of confidence in the overall clinical comparison: “Since tirzepatide and empagliflozin have completely different mechanisms of action, without a direct comparison, it is difficult to judge whether tirzepatide may represent a substantial improvement over empagliflozin, particularly	

	in patients with established or at high risk of ASCVD, CKD, or heart failure; three common co-morbid conditions". Due to lack of evidence on CV and renal outcomes for tirzepatide and lack of consideration for corresponding data available for empagliflozin, the evidence base for this clinical assessment is incomplete and does not allow for a definitive rating of tirzepatide's net health benefit compared to empagliflozin. BI recognizes that ICER acknowledges the lack of long-term evidence on cardiorenal metabolic effects of tirzepatide, but urges ICER to reflect the considerable uncertainty inherent in this assessment by revising its comparative evidence rating to "I".	

5.	2. Concerns related to ICER's overall modeling approach In addition to our comments on the comparative effectiveness rating for tirzepatide compared to empagliflozin, BI would like to point-out several concerns regarding ICER's overall cost effectiveness (CE) modeling approach. BI commends ICER for providing a version of the CE model for review, but would like to highlight three important concerns that lead to high model and parameter uncertainties and limit our confidence in the results: A. UKPDS OM2 risk engine is not well-suited to represent current treatments for T2DM B. Assumptions regarding treatment discontinuation are not reflective of clinical practice C. The model does not adequately represent empagliflozin's adverse event rates observed in clinical trials. A. The UKPDS risk engine does not reflect cardiorenal metabolic aspects of T2DM and does not represent current population dynamics The initial UKPDS population is based on newly diagnosed T2DM patients in the UK from 1977 - 1997. This population fundamentally differs from ICER's US-based target population with respect to demographic and health characteristics, available medications and dietary preferences. Moreover, diagnosis and treatment patterns have evolved substantially over the past 20 years, which likely have changed underlying risk relationships described in the UKPDS OM2. The UKPDS includes 5,102 newly diagnosed patients with T2DM, and risk equations derived for this cohort are not representative of the risk of CV and renal events for patient populations from CVOTs with an average T2DM duration of over 10 years. CVOTs like EMPA-REG OUTCOME enrolled around 7,000 patients, with an average follow-up of more than three years. Risk equations derived for patients at high risk of CV events will yield greater accuracy in projection of CV and renal events. Thus, they should be used in a CE analysis for patients with increased CV risk or prevalent CKD, instead of UKPDS. A model relying on the UKPDS risk equations will not represent the benefit of a ketogenic state, reductions in glomerular pressures with preservation of renal function, and lower left ventricular filling pressures that are independent of BP lowering and volume contraction. It is these pleiotropic effects that mostly touch the comorbidities that account for the greatest morbidity and healthcare utilization of patients	We have augmented our discussion of the limitations of the UKPDS OM2 in the Evidence Report. We also performed two sets of additional simulations with the model using time horizons that aligned with the CVOTs for injectable semaglutide and empagliflozin. In Supplement Table E6 we present comparisons of the model's predicted MACE, CV mortality, and all-cause mortality to those events observed in the CVOTs, for semaglutide, empagliflozin, and placebo (compared to background therapy alone model predictions). We note that the model predicted slightly higher rates of MACE composite events, but slightly lower CV mortality rates, slightly lower all-cause mortality in the 2-year simulation and slightly higher in the 3-year simulation. However, the comparison between trial outcomes and model outcomes overall is of a similar scale.

	with T2DM, yet are unaccounted for in available risk engines such as OM2 or BRAVO. To further illustrate this shortcoming of the UKPDS OM2 risk engine, a simulation of the OM2 with the EMPA-REG OUTCOME data revealed that the OM2 only accounted for 12.75–15% of the overall CV benefit of empagliflozin. The documented limitations of available T2DM risk engines such as BRAVO and the OM2 in representing benefits of SGLT2 class, introduce substantial model-based uncertainty to ICER's assessment, on top of the aforementioned data-based uncertainty inherent in the indirect comparison approach. To reflect this magnified level of model uncertainty, BI reiterates its request that ICER revise the evidence rating to "I", and to state explicitly in the main text of the evaluation the documented shortcomings of the UKPDS OM2 in accounting for the CV benefit of empagliflozin.	

6.	B. Assumption around treatment discontinuation in ICER's model does not reflect clinical practice guidelines In the model, treatment discontinuation occurs if HbA1c exceeds 8.5% (see Table 4.1). In clinical practice, patients with HbA1c exceeding 8.5% would receive additional glucose lowering agent and not discontinue their SGLT2 inhibitor. The 2021 ADA guidelines recommend SGLT2 inhibitors be continued for cardio-renal protection, irrespective of how effective they are for patients achieving their HbA1c goal. In the 2019 T2DM evaluation, ICER assumed that following the first model cycle, "oral semaglutide, empagliflozin, and liraglutide patients added insulin therapy while remaining on their current treatment if their HbA1c reached 8.5 or above". EMPA-REG EXTEND is referenced as ICER's central data source for treatment discontinuation of each treatment under review. However, this trial was a safety extension of the EMPA-REG 26 week clinical trial which, in order to observe patients with a longer exposure without rescue, introduced a discontinuation mechanism as a safety precaution for patients not achieving goal of HbA1c ≤8.5%. This threshold is not reflected in the ADA guideline issue and was specifically introduced for the conduct of this FDA-mandated safety study to increase exposure of at least 400 patients to 2 years of exposure. The discontinuation rate in EMPA-REG EXTEND should therefore not be used to mimic real world use of empagliflozin. BI recommends that ICER's model consider continuous use of treatment, rather than discontinuation based on HbA1c levels, as was done in the 2019 T2DM assessment of oral semaglutide	Thank you for this suggestion. We have updated the model to continue active treatments for the patient's lifetime and instead add on insulin when the patient's HbA1c exceeds the specified threshold.
7.	C. Adverse event rates are not representative of empagliflozin's clinical trial data, even after ICER's adjustments for CV event rates ICER notes "because no long-term cardiovascular outcomes trial data exist for tirzepatide, health benefits were informed by intermediate outcomes and were unadjusted. Modeled cardiovascular and renal outcomes for therapies with existing long-term trials were adjusted to trial data using hazard ratios." ICER used inputs from the NMA for efficacy at reducing HbA1c, weight, SBP, and LDL for all treatments, and then applied event reduction hazard ratios from the CVOTs (in addition to benefits treatments garnered from reductions in intermediate outcomes) to both empagliflozin and semaglutide. The	See above in Item 6 (point A) for our description of additional analyses reported in the Evidence Report that capture shorter time horizon outcomes aligned with the CVOTs.

	incidence rates of key cardiorenal metabolic outcomes such as CHF, composite MACE, CV death and renal death, observed for patients treated with empagliflozin in EMPA-REG OUTCOME differed substantially from outcomes projected in ICER's model. Even after ICER's calibration, adverse event rates for empagliflozin are overestimated compared to its published data, which, in consequence, leads to an underestimation of key benefits in the model, including LYs and QALYs. Additionally, it remains unclear how various aspects of the comparison such as HbA1c and weight loss are weighted relative to other model inputs, thereby operating as a "black box". While ICER performed a calibration exercise, the calibration process and end results are lacking in both clarity and transparency. BI recommends ICER's calibration more closely align with EMPA-REG OUTCOME data, in order to adequately represent the full range of value that empagliflozin provides for T2DM patients.	
8.	Additional analyses would enhance ICER's model and provide clarity around model assumptions: BI recommends including a scenario of life-long treatment, given the CV benefit of EMPA irrespective of HbA1c and ADA guidelines. See discussion above within treatment discontinuation. ICER assumption of a constant BMI post-treatment discontinuation impacts LY and QALYs. ICER should model the impact of this assumption on outcomes, including patient's BMI reverting to the original level, post treatment discontinuation. For model transparency, BI recommends including the calculations underlying the model either in the report or with the model when delivered.	Thank you for these suggestions. The Evidence Report includes a base case where treatment is life-long other than the initial risk of all cause discontinuation. We also added scenario analyses where we turned off risk factor progression to maintain the impact on BMI and HbA1c, the results of which may be found in the supplement. The equations utilized in the model are all derived from publicly available manuscripts that are cited in the Evidence Report.

Eli Lilly		
1.	ICER's base-case analysis should use a cardiovascular (CV) event hazard ratio (HR) for tirzepatide from either the SURPASS-4 clinical trial or the CV safety meta-analysis of tirzepatide clinical trials. ICER should also conduct a scenario analysis with NO adjustment for CV outcomes using CV event HRs. In the draft evidence report, ICER applies an adjustment for semaglutide and empagliflozin based on the CV event HRs from their CV outcome trials (CVOTs) but assumes a CV event HR of 1.0 (i.e., no adjusted CV benefit) for tirzepatide. This is a flawed assumption as there is early evidence suggesting that tirzepatide has a potential CV benefit and contradicts International Society for Pharmacoeconomics and Outcomes Research (ISPOR)'s good research practices as reported by the ISPOR Modeling Good Research Practices Task Force (Briggs 2012, Caro 2012, Eddy 2012). By applying adjustments to semaglutide and empagliflozin but not to tirzepatide, ICER is creating an uneven comparison by potentially double-counting benefit for semaglutide and empagliflozin and assuming no adjusted benefit for tirzepatide. Lilly recommends revising the base-case analysis so that an empirically supported adjustment is applied to tirzepatide's estimated CV outcomes. Given that ICER acknowledges that "a meta-analysis of cardiovascular events for safety across the SURPASS trials showed no increase in cardiovascular events and a trend toward cardiovascular benefit," the base-case CV event HR should reflect the best current estimate of the potential benefit for tirzepatide. ICER should use a HR for tirzepatide from either the SURPASS-4 clinical trial or the CV safety meta-analysis of tirzepatide clinical trials. Although tirzepatide's CVOT is currently ongoing, peer-reviewed data from SURPASS-4 (which enrolled a high-risk CV population) is available (Del Prato 2021). Results found adjudicated 4-point major adverse CV events (MACE-4; CV death, myocardial infarction, stroke, hospitalization for unstable angina) occurred in 109 participants and were not increased on tirzepatide compared with insulin glargine. Lilly also conducted a CV safety meta-analysis across the tirzepatide clinical program once the predefined number of MACE occurred (Lilly 2021). The meta-analysis consisted of 116 participants with adjudicated CV events contributing to the MACE-4 outcome, the majority of which came from SURPASS-4 (Lilly 2021). This recommended base-case analysis should be conducted not only for comparison of tirzepatide vs	We have updated the model and the Evidence Report so that the base case results utilize the MACE HR and its uncertainty from SURPASS-4. When selecting between the meta-analysis and SURPASS-4 in terms of MACE HR, we slightly favored the HR from SURPASS-4 as the vast majority of the data in the meta-analysis comes from SURPASS-4 and SURPASS-4 is peer-reviewed. We included the assumed CV event HR of 1.0 as a scenario analysis, reported in the supplement.

	background therapy but also comparison of tirzepatide vs active comparators (semaglutide and empagliflozin). Additionally, ICER should conduct a scenario analysis with no additional adjustment for CV outcomes for tirzepatide, semaglutide, and empagliflozin based on CV event HRs. Because the UKPDS risk equations are intended to model CV outcomes without adjustment using CV event HRs, applying a CV event HR on top of the risk equations could result in double-counting of CV benefit. Therefore, it is important to include a scenario analysis with no HRs for any treatments to illustrate the impact of adjustments to the model.	
2.	The uncertainties and limitations of the cost-effectiveness analysis when interpreting the results and discussing conclusions should be clearly described in the report. ICER acknowledges that there is a "wide range of plausible cost-effectiveness estimates for tirzepatide" (ES3), with cost-effectiveness ratios for tirzepatide vs semaglutide ranging from -$1,469,000/quality-adjusted life-year (QALY) to $1,541,000/QALY and for tirzepatide vs empagliflozin ranging from -$408,000/QALY to $594,000/QALY. The significant overlap in the credible intervals for tirzepatide and semaglutide in costs and QALYs indicate that there is not a conclusive difference in cost-effectiveness between the 2 drugs, as small changes in costs or QALYs could completely change the cost-effectiveness ratios. Given the uncertainties in the model results due to wide ranges of estimates, ICER should make it very clear that results and conclusions are based on many uncertainties and assumptions when discussing the interpretation and conclusions of the cost-effectiveness analysis. Making conclusive statements regarding cost-effectiveness without acknowledging the limitations and uncertainties of the analysis could impact access to valuable treatments for T2D patients.	We have updated the conclusion statements to more accurately reflect adjustments made to the modeling approach and uncertainty in the results.
3.	As Tirzepatide is not currently approved and does not have a published price, Lilly recommends a threshold analysis of tirzepatide that uses the assumption of price parity to semaglutide 1.0mg to determine the HR required to reach cost-effectiveness to each of the agents. ICER should also conduct a threshold analysis of tirzepatide that uses the assumption of price parity to semaglutide 1.0mg to determine the CV event HR required to reach cost-effectiveness compared to semaglutide, empagliflozin, and background therapy. Given that the price of tirzepatide is currently unknown and there is a level of uncertainty on the long-term CV outcomes of tirzepatide, conducting a threshold analysis to determine the HR required to reach cost-effectiveness for tirzepatide would provide readers	The goal of our analysis is to provide value-based threshold prices. We have updated the base case to utilize the CV event HR from SURPASS-4 rather than conduct a threshold analysis on tirzepatide's effectiveness.

	more information about the cost-effectiveness of tirzepatide once there is pricing available and the long-term CVOT for tirzepatide is completed.	
4.	ICER should include the difference in device preference utilities between tirzepatide and semaglutide in a scenario analysis of the cost-effectiveness model. Patient preference is an important consideration when choosing a diabetes treatment, as route of administration, frequency of administration, and injection device can affect adherence and quality of life. Lilly was pleased that ICER acknowledged differences in patient preferences for diabetes treatment by including an annual disutility for daily injection of insulin (for patients who discontinued treatment) based on a publication by Boye et al. (2011). In addition to an injection disutility for insulin (a non-active comparator), ICER should consider including quality of life data for active comparators as well, for which there is recent published and peer-reviewed data. There is a well-established difference in device preference between the injection devices for semaglutide and tirzepatide (which is the same as the dulaglutide injection device). Results from a recent study (Boye 2019) of the semaglutide and dulaglutide injection devices showed a mean (SD) utility difference between the injection device health states of 0.007 (0.019). ICER should include this difference in device preference utility between the tirzepatide and semaglutide devices in a scenario analysis of tirzepatide compared to semaglutide	We include a disutility for insulin due to its daily administration, but do not include one for other active injectable treatments (tirzepatide or semaglutide) due to the lower frequency of injection. As such, the modeling is not providing an advantage/disadvantage for any one treatment. However, we do acknowledge that device preference may be important in the potential other benefits section of the Evidence Report.
5.	Since ICER used the UKPDS model rather than the BRAVO model despite known limitations, more detail should be provided on its limitations, the impact these limitations have on the interpretation and accuracy of the model outcomes, and the process followed to select UKPDS OM2 for use in the assessment model over other non-BRAVO risk engines. ICER indicated that the BRAVO risk engine would be used in the model analysis plan, so more details should be provided on the difficulties implementing the BRAVO risk equations, what additional models were considered when BRAVO was deemed infeasible for use in the assessment, and why the best alternative was then determined to be the UKPDS OM2 given the known limitations of UKPDS-OM2. ICER states that "the UKPDS-OM2 risk equations are widely used in diabetes simulation models and have been shown to accurately predict results for the population in which it was developed as well as other diabetes populations" (page E4). While ICER is correct that the UKPDS-OM2 risk equations predict results for the United Kingdom (UK) population as well as populations similar to the UK (e.g., Ireland, Scotland, etc.), it	We have augmented our discussion of the limitations of the UKPDS OM2 in the Evidence Report. Additionally, we performed two sets of additional simulations with the model using time horizons that aligned with the CVOTs for injectable semaglutide and empagliflozin. In Supplement Table E6 we present comparisons of the model's predicted MACE, CV mortality, and all-cause mortality to those events observed in the CVOTs, for semaglutide, empagliflozin, and placebo (compared to background therapy alone model predictions). We note that the model predicted slightly higher rates of MACE composite events, but slightly lower CV mortality rates, slightly lower all-cause mortality in the 2-year simulation and slightly higher in the 3-year simulation. However, the comparison

	has not been shown to accurately predict results among United States (US) populations. Evidence from the last two Mount Hood Meetings provide little evidence to support the use of the UKPDS OM2 in the US due to differences in racial and ethnic characteristics between the 2 populations, along with differences in diabetes characteristics, such as a great proportion of obese individuals, higher body mass index (BMI), younger age of diagnosis and diagnosis at a lower baseline HbA1c level for US patients (Palmer 2018, Si 2020). There is a paucity of published validation evidence supporting the use of UKPDS OM2 risk equations in populations taking newer treatments for type 2 diabetes (T2D), particularly for interventions associated with weight loss. Evidence from Mount Hood indicated that the risk equations needed re-calibration to provide plausible estimates of outcomes from CVOTs (Palmer 2018). Additionally, UKPDS OM2 is based on patient-level data for T2D patients who were recruited between 1977 and 1991 and were followed until 1997 (Hayes 2013), making this a very outdated population. ICER should provide additional details on the limitations of the UKPDS OM2, including that it is outdated and not validated in a US population, and that the risk equations have been demonstrated to poorly predict CV outcomes, as well as the impact these limitations have on the interpretation and accuracy of the model outcomes	between trial outcomes and model outcomes overall is of a similar scale.
6.	Given the considerable uncertainty in results, ICER should conduct additional sensitivity analyses on key parameters driving model uncertainty. Given this considerable uncertainty in results described above, ICER should conduct additional sensitivity analyses for the revised evidence report, including conducting sensitivity analyses on risk factor progression assumptions, device utility, and different weight gain utility approaches, different QALY estimation approaches, and CV event HRs. Any of these factors could considerably change cost-effectiveness results, so it is important to demonstrate the impact of each factor on these results	Thank you for these suggestions. We have added scenario analyses in the Evidence Report that include turning off the risk factor progression equations, eliminating the disutility for obesity, and differences in HRs applied to the risk equations.
7.	ICER indicates that efficacy inputs were derived from the network meta-analysis (NMA); however, given that there is head-to-head trial data for tirzepatide vs semaglutide from SURPASS-2, ICER should use these inputs instead or provide clear rationale for why the NMA-derived data were deemed more appropriate for the comparison of tirzepatide to semaglutide than the SURPASS-2 data provided by Lilly. On page 23, ICER states that the "effects of each included therapy, such as change in HbA1c after the first cycle, were included depending on data availability from the NMA."	We are utilizing NMA outputs for risk factors in order to have a consistent approach for all comparisons in the model, with background therapy alone as the common thread. We also note that the point estimate/uncertainty from the NMA matches up with the head-to-head clinical trial for intermediate endpoints.

	ICER also states on page 27, "clinical inputs regarding the efficacy of tirzepatide, injectable semaglutide, and empagliflozin as compared to background therapy alone on intermediate outcomes such as changes in HbA1c, lipid levels, blood pressure, and body weight were derived from the NMA described in Chapter 3." ICER should use efficacy inputs from the head-to-head (SURPASS-2) trial data for tirzepatide and semaglutide instead of the data from the NMA, as direct head-to-head data is preferred to indirect treatment comparisons (such as from an NMA) in the hierarchy of strength of evidence. ICER should also provide a clear rationale for why the NMA-derived data were deemed more appropriate for the comparison of tirzepatide to semaglutide than the SURPASS-2 data provided by Lilly	
8.	Lilly recommends that ICER quality check the data inputs in their NMA and economic model to ensure the use of the tirzepatide 15 mg dose data. On Page 11, ICER comments that there is a mean HbA1c difference in tirzepatide from background therapy of -1.7% from the NMA. In Table D2.2, the data inputs for the NMA report the change from baseline from the SURPASS-2 study (tirzepatide = -2.3%; semaglutide = -1.86%) and HARMONY-3 (background therapy = 0%; sitagliptin = -0.5%). Table 2.4 shows the results of the NMA and reports a difference of -1.72% between tirzepatide and background therapy. When Lilly ran an NMA using the same inputs, there was a difference in HbA1c of -2.0% between tirzepatide and background therapy. If the NMA is re-run using the 5 mg result for tirzepatide (-2.01%), the outcome from the NMA for tirzepatide vs background therapy matches the value reported in Table D2.4 (-1.7%). We believe that the NMA for HbA1c has incorrectly used the 5 mg tirzepatide result as the input instead of using the 15mg tirzepatide result. We recommend that ICER quality check the data inputs in their NMA and economic model to ensure the use of the tirzepatide 15 mg dose data. Similar changes will need to be made to the difference between tirzepatide and semaglutide or empagliflozin if the 5 mg result has been incorrectly used in the NMA.	We appreciate the check on our data inputs in the NMA. We have reviewed the data in Table D2.2, and they match what was provided to ICER by Lilly for tirzepatide 15 mg, and also what is publicly available in the SURPASS-2 publication. Discrepancies in the NMA results may be due to values that were provided to us by other manufacturers as academic-in-confidence data.
9.	Tirzepatide's clinical evidence rating vs semaglutide should be changed from a C+ to a B+ or C++. Tirzepatide was given a comparative clinical effectiveness grade of C+ (comparable or incremental) in comparison to semaglutide despite substantial improvements in nearly all outcomes of interest. Tirzepatide showed an improvement of 0.45% in HbA1c, an additional 5.5 kg weight loss, and an additional decrease of 2.9 mmHg in systolic blood pressure	As we have stated in the report, we agree that tirzepatide provides superior improvements in HbA1c, weight, and systolic blood pressure compared with semaglutide. However, semaglutide has demonstrated long-term cardiovascular benefit, while tirzepatide's formal cardiovascular outcomes trial is ongoing.

	in comparison to semaglutide. Despite the lack of long-term CV outcomes, early intermediate outcomes suggest a trend toward a strong CV benefit. This was acknowledged by ICER as well. As a result, ICER should consider changing tirzepatide's clinical evidence rating from C+ (comparable or incremental) to a B+ (incremental or better) or C++ (comparable or better) in comparison to semaglutide.	Although SURPASS-4 suggests potential cardiovascular benefit, the results were not statistically significant and thus cannot be taken as definitive. A rating of C++ is used when there is greater uncertainty about benefits, for example, when comparisons are indirect such as those between tirzepatide and empagliflozin. Because there was a head-to-head trial with semaglutide, we have more certainty about the relative benefits of tirzepatide compared with semaglutide and thus are comfortable with the C+ evidence rating.
10.	In addition to the above recommendations, some additional information/data is needed to interpret the results of the cost-effectiveness model, including: • Additional information from the NMA output, including 95% credible intervals and results from the different models run, along with deviance information criterion (DIC) values and residual deviance values so that readers may assess the models' goodness of fit data	In our revised Evidence Report supplement, Table D2.4 provides the point estimates and 95% credible intervals for Tirzepatide vs Background Therapy and Tirzepatide vs Empagliflozin. Our Evidence Report relies on only one model, a random effects model with informative priors. Our reasoning for selecting this model over the initial two models is provided in Supplement D2.
11.	• Results of ICER's model validation where ICER varied the model input parameters to evaluate the face validity of changes to those inputs on the results • Clinical event rates and risk factor progression over time to aid in interpretation of cost-effectiveness results	As discussed in a previous comment, we performed two sets of additional simulations with the model using time horizons that aligned with the CVOTs for injectable semaglutide and empagliflozin as a form of external validation. In Supplement Table E6 we present comparisons of the model's predicted MACE, CV mortality, and all-cause mortality to those events observed in the CVOTs, for semaglutide, empagliflozin, and placebo (compared to background therapy alone model predictions). The overall comparison between trial outcomes and model outcomes is of a similar scale. While the nature of the micro-simulation model does not easily allow for clinical event rates and risk factor progression output over all possible times, the shorter

		time horizon scenarios offer insight into how the model predicts event rates in a time frame comparable to CVOTs.
12.	• State diagrams to allow readers to see patient progression across the different model comparators	The micro-simulation does not output patient-level data or temporal data, but rather aggregate data over the stated time horizon. While a state diagram would therefore be inappropriate, we do outline our model structure in Figure 4.1 of our Evidence Report.
13.	• Rationale for using an additive approach to combining QALY disutilities when the coefficients described by Shao et al. (2019) were designed to be combined in a regression formula (eg, OLS regression) • Scenario analyses exploring the impact of using an additive approach vs regression formula for QALY disutilities	The model approach to utilities uses an OLS regression approach, taking an intercept and adding disutility betas where factors are present for a patient at a given time in the model. The language around this approach has been clarified in the report.
14.	• Full disaggregated results with costs and outcomes stratified across all available categories (eg, AEs, CV outcomes, renal outcomes, insulin, etc.) to help assess what is occurring in the model per modeling best practices and many health economics and outcomes research guidelines from around the world	Disaggregated results have been added to the supplement in table E3.1.
15.	• One-way sensitivity analysis (OWSA) results for tirzepatide compared to semaglutide and empagliflozin	Additional OWSA tornado plots comparing tirzepatide to additional comparators have been added to the supplement section E4.
16.	• Undiscounted results from the cost-effectiveness analysis so that the budget impact analysis can be validated	Undiscounted results have been added to the supplement in table E3.2.
17.	Furthermore, there are several areas that lack sufficient information to evaluate and replicate (for those with modeling expertise) ICER's cost-effectiveness model. ICER should include clear details in the report on the following areas: • How treatment discontinuation is applied throughout the model • How HRs are being applied in the model to adjust the CV and renal outcomes • Clinical inputs and the risk factor progression • The report indicates that time varying values of HbA1c and weight were calculated using	Thank you for these thoughtful suggestions. The Evidence Report includes extensive updates responding to these points. Specifically responding to Bullets 6-7: The OWSA was re-run to incorporate other model changes, and this no longer appears to be an issue.

	additional published equations from Willis 2017, but this publication only provides regression functions for changes in HbA1c and body weight on insulin initiation and does not provide estimates of risk factor progression over time. Moreover, information on insulin doses and insulin types are needed for these equations described in Willis 2017 and these are not provided. • Risk factor progression is critical in terms of understanding the analysis, particularly with respect to the timing and impact of discontinuation, HbA1c difference between treatment arms, and the influence of BMI over time on quality of life. • Explanation for why high and low parameter estimates both results in higher incremental QALYs than the implied base-case value when parameters are varied in Figure 4.3 • For example, varying the tirzepatide HR for nephropathy yields incremental QALYs of approximately 0.68 for the high estimate and 0.62 for the low estimate, whereas the base-case QALY appears to be around 0.59. This anomaly applies to the majority of the parameters. It should be clearly explained how and why both the low and high estimates would be higher than the base-case estimate. • Limitations of using the NHANES cohort for ICER's base-case model cohort • Approximately 32 million people in the US have T2D (Dugani 2021). However, ICER used a cohort of only 387 patients, which is a small sample of patients to be representative of the entire US T2D population. • The proportion of smokers in the US general population has been estimated at around 14% (Cornelius 2020), so ICER's estimate of 36.7% in the base-case model cohort seems high. • Similarly, estimates of concomitant medication use (100% on metformin and 42.9% on sulfonylurea) seem high. • Additionally, "renal disease" appears to be costed in Table E.4 as end-stage renal disease (ESRD), but it seems implausible that 15.8% of the population has ESRD at baseline. For comparison, the	

	population in the Yang 2020 population (used for costing) has this estimate at 0.54%.	
18.	Given ICER's commitment to open and transparent engagement with stakeholders in their reviews, ICER should allow stakeholders the ability to provide input on results and analyses that were not presented as part of the draft evidence report.	We agree that stakeholder input throughout each stage of our process is important, including the four-week public comment period on our draft evidence report, and make updates accordingly before publishing our evidence report. We also include all oral public comments delivered at our public meeting after the evidence report is published.
Novo Nordisk		
1.	We would like to reinforce our agreement with ICER on the following assumptions and choices: I. No established benefit for tirzepatide for cardiovascular outcomes Novo Nordisk agrees with ICER that it is appropriate to not assume additional benefit for tirzepatide on cardiovascular outcomes given that there is currently no data for tirzepatide from a cardiovascular outcomes trial.	Thank you for this feedback. As mentioned in other responses, we received feedback in both directions on this economic model assumption and choice. In the evidence report, we decided the hazard ratio (including its uncertainty) from SURPASS 4 was the best-available evidence for tirzepatide on cardiovascular outcomes.
2.	II. Insufficient data for patients with comorbid CKD to evaluate at this time Novo Nordisk agrees with ICER that patients with type 2 diabetes and comorbid chronic kidney disease (CKD) represent an important patient population, but data is currently insufficient to conduct an evaluation in this population at this time. The FLOW trial was initiated to explore the impact of semaglutide in patients with type 2 diabetes and CKD, with results expected in 2024.	Thank you for this feedback!
3.	We would like to provide the following suggestions that we believe will improve the findings and make the report more useful to stakeholders: I. Placeholder net price for tirzepatide likely underestimates actual net price at launch The ICER price estimate for tirzepatide is likely inaccurate and will limit the usefulness of cost-effectiveness findings for payers and other stakeholders. The assumed price as equivalent to semaglutide is unlikely to represent the net price at launch. To rectify this, we suggest using the SSR database to assess rebates *at launch* from other GLP-1 products, and correspondingly adjusting the rebate percentage suggested for tirzepatide in the model to make the model representative of what will most likely happen in the real world.	Given that tirzepatide's price is a placeholder price, the Evidence Report emphasizes health benefit price benchmarks (which are unrelated to placeholder price) as opposed to incremental cost-effectiveness ratios for tirzepatide. We also emphasize that results based on a placeholder price should be interpreted with caution throughout the Evidence Report.

4.	II. Influence of serious adverse events and discontinuation due to adverse events on model outputs is unclear Although no single serious treatment related adverse events occurred in either treatment arm in ≥5% of patients (ICER's threshold) in the SURPASS-2 trial, the overall rate of serious adverse events was higher with tirzepatide 15 mg (5.7%) vs semaglutide 1 mg (2.8%). In addition, discontinuations due to adverse events were approximately double with tirzepatide 15 mg (8.5% vs 4.1%). We look forward to additional clarity on how serious adverse events and adverse events leading to discontinuation are considered within in the model, given the direct head-to-head data from the SURPASS-2 trial.	We accounted for all cause discontinuation after the first model cycle to handle AEs and all other reasons for discontinuing treatment as there were not specific individual events to directly model. We do not anticipate that AEs would be a significant driver of lifetime cost-effectiveness because not any one rises to 5% or more.
5.	III. Long-term data is suggestive of waning glycemic durability of response for some agents If the team hasn't already considered longer term trials, such as the EMPA-REG OUTCOME trial, we encourage the inclusion of data pertaining to the waning impact on glycemic control over the study duration. We feel there may be implications for specific model outcomes such as the proportion of patients at any glycemic control threshold. Based on these studies and lack of long-term comparative head-to-head data, there is considerable uncertainty around the comparative durability of response.	Thank you. At this time, we do not have longer term data for tirzepatide, the main focus of this report. We relied on the assumption that the natural history of diabetes involves risk factor progression.

#	Comment	ICER Response
Researchers and Economists		
Ossian Consulting		
1.	Unfortunately, the reporting in the Long-term Cost-effectiveness section of the ICER Draft Evidence Report falls well short of the recommendations outlined in these publications; a shortcoming which may lead to stakeholders questioning the credibility of the modeling analysis. Moreover, it would be impossible for independent researchers to reproduce the analysis (despite it being based on a published model) as several key aspects of the modeling analysis are inadequately described or missing altogether from the report. As noted by the Mount Hood authors, reproducibility is likely to enhance the credibility of any modeling analysis.	Thank you for this suggestion. We have augmented the methods section in both the Evidence Report and Supplement to provide further details about the modeling methods.
2.	Transparency of Methodology There are several notable omissions from the Methods section of the report, including a clear description of initial treatment effects on baseline biomarkers (the reference to Chapter 3 does not afford the reader any clarity on precisely which values were included in the modeling analysis). The same criticism can be leveled (either in terms of lack of clarity or complete omission) at the description of any of the following aspects of treatment taken from the Mount Hood checklist: Trajectory of biomarkers, BMI, smoking, and any other factors that are affected by treatment • Treatment algorithm for HbA1c evolution over time • Treatment algorithm for other conditions (e.g., hypertension, dyslipidemia, and excess weight) • Rules for treatment intensification (conflicting descriptions are provided in the body of the report and in the supplementary material) • Long-term effects, adverse effects, treatment adherence and persistence, and residual effects after the discontinuation of the treatment	See above. We provide more details in the report supplement.
3.	In addition, there are several technical aspects of the modeling analysis that are missing from the report. For example, cohort characteristics are summarized in Table 4.2 on page 27 of the report but no information is provided on how the race categories described were reconciled with the ethnic groups employed in the United Kingdom Prospective Diabetes Model Outcomes Model 2 (UKPDS OM2) risk equations, which were used to evaluate the risk of complications and mortality in the model. Descriptions of	See above. We provide more details in the report supplement. We have also updated the report to be more explicit about the use of HRs to adjust the model event predictions.

	the distributions used around model parameters in probabilistic sensitivity analysis (PSA) are not provided, despite all base case and sensitivity analysis results being derived from analyses in which PSA was active. Perhaps more critical is the lack of detail regarding the implementation of hazard ratios to adjust the risk of diabetes-related complications for each intervention based on data from cardiovascular outcomes trials (CVOTs) (or assumption in the case of tirzepatide).	
4.	Whilst the report is clear (Table E.2 on page E3) on the hazard ratios for the risk of major adverse cardiovascular events (MACE), congestive heart failure and nephropathy (which we assume to correspond to renal failure in the modeling analysis), how these hazard ratios are applied in the modeling analysis is simply not described. This is a critical feature of the modeling analysis as the data presented in the report indicate that it is a key driver of outcomes. MACE is a composite endpoint that is not evaluated by the UKPDS OM2 risk equations. And whilst there are UKPDS OM2 risk equations that can be used to determine which simulated patients experience myocardial infarction and/or stroke events, there is no risk equation that is specific to cardiovascular death, typically the third endpoint included in the three-point MACE definition. We would suggest that it is critical to the credibility of the modeling analysis that transparency is improved in this area. It is perhaps worth noting that we have restricted our comments here to the issue of transparency and have left aside the serious limitations that may be associated with applying unadjusted hazard ratios to modeled endpoints in the analysis.	See above. We provide more details in the report supplement. We have also updated the report to be more explicit about the use of HRs to adjust the model event predictions. Additionally, we provided shorter time horizon simulations in the supplement where we compare the model outputs to the outcomes observed in the empagliflozin and injectable semaglutide CVOTs.
5.	Transparency of Results Similarly, reporting of results is inadequate. We would suggest that, at a minimum, the report needs to include survival curves, descriptions of the incidence of diabetes-related complications over time for each of the interventions and a breakdown of costs for each simulation arm. This would allow readers to better understand how the changes in risk factors associated with the different interventions in this review influence complication rates, and the role that hazard ratios play in adjusting the complication rates predicted by UKPDS OM2 equations. For all PSA simulations, we would recommend the presentation of scatter plots of incremental costs versus incremental quality-adjusted life years (QALYs) to inform the reader on uncertainty around the reported outcomes.	Thank you for these suggestions. We added many of these suggested details in the Evidence Report and supplement.

6.	Justification of Assumptions In the report, only four of the base case assumptions are justified (see Table 4.1 for details) and we would suggest none of these would be counted as critical base case assumptions. This falls short of what would be expected, for example, in any manufacturer's submission for health technology assessment in other countries and, most likely, what would be needed to successfully publish in a peer-reviewed journal. Key justifications are needed around the choice of model risk equations (UKPDS OM2 risk equations for the US setting); at present only implementation challenges are cited as the reason for choosing UKPDS OM2 equations for the model. The authors could also explore the likely impact of this choice by employing other published risk equations in sensitivity analyses. Similarly, assumptions around risk factor progressions, triggers of treatment intensification, the additive approach to estimating QALYs (ignoring published regression formulae based on the same data), body mass index- and treatment device-related utilities, and the use of hazard ratios to adjust the risk of complications in the base case are not justified in the report. Finally, the choice of cohort should be justified given that the 387 patients selected from NHANES (Table 4.2) do not appear to be representative of the type 2 diabetes population in the US in certain respects.	We have addressed many of these points in responses to other comments. The Evidence Report expands on the model's assumptions, rationale, and methods.
7.	Improving transparency To ensure the credibility of the cost-effectiveness analysis in ICER's ongoing review of tirzepatide in type 2 diabetes, we would encourage the reviewers to follow the recommendations for transparency laid out by the Mount Hood group in 2018. This would produce an expanded report but having a transparent, reproducible modeling analysis would greatly enhance the credibility of the cost-effectiveness evaluation and acceptance of ICER's findings by almost all key stakeholders. ICER should go further and share the cost-effectiveness model in Microsoft Excel beyond the small group of stakeholders currently afforded access for review. A version of the model with all commercial-in-confidence data removed that could be reviewed by a larger group of interested parties would do much to improve its validation. Further, ICER should share a version of the model without password protection to all stakeholders. So far as can be ascertained from the report, there is little substantive intellectual content in the model that should be proprietary to ICER. There should, therefore, be no barrier to sharing the model with a wider review group; a step which would improve transparency and validation as well as building confidence in the quality of the model implementation.	We agree that transparency about our model is important, and our standard approach is to include enough detail between the evidence report and the supplement for someone with health economic training to replicate our work. The rights to the model belong to the University of Washington, but we have created a model transparency program to share executable versions of our draft cost-effectiveness models with relevant drug manufacturers during our public comment period. Of note, two manufacturers participated in the model transparency program for this review.

8.	We would also suggest that an appendix be prepared that describes the validation analysis (with input settings and results) for inclusion in the overall report. Given the choice of risk equations for the model and the approach of applying unadjusted hazard ratios to model outcomes, validation analyses against CVOTs as well as in type 2 diabetes populations comparable to the US population are needed to support the existing model.	We performed two sets of additional simulations with the model using time horizons that aligned with the CVOTs for injectable semaglutide and empagliflozin. In Supplement Table E6 we present comparisons of the model's predicted MACE, CV mortality, and all-cause mortality to those events observed in the CVOTs, for semaglutide, empagliflozin, and placebo (compared to background therapy alone model predictions). We note that the model predicted slightly higher rates of MACE composite events, but slightly lower CV mortality rates, slightly lower all-cause mortality in the 2-year simulation and slightly higher in the 3-year simulation. However, the comparison between trial outcomes and model outcomes overall is of a similar scale.
Tulane University		
1.	We re-assessed the cost-effectiveness analysis of GLP-1 agents in a published systematic literature review. Among a total of 48 CEA studies using the diabetes models based on the UKPDS risk engine, we found that better long-term effectiveness results were driven by better improvements in biomarkers (HbA1c, LDL, and BP) in 47 CEA studies. Because the evidence report also used the UKPDS, it is plausible that long-term effectiveness analysis of tirzepatide versus semaglutide would favor tirzepatide. However, the tirzepatide was less effective in QALY (Table 4.5) in the base case scenario. In addition, we are puzzled by the methods of applying hazard ratios (Table E.2. page E3) for tirzepatide, semaglutide, and empagliflozin in the draft evidence report. We are not clear about how the hazard ratios were applied into the UPPDS OM2, and how the detailed processes of long-term effectiveness were derived.	We have revised the Evidence Report extensively, including adding sentences clarifying how risk reductions were applied using HRs multiplied by the UKPDS event predictions in each model cycle. We also note, as mentioned above, that the base case has been revised to apply a HR for tirzepatide's cardiovascular outcomes using data from the SURPASS-4 trial.
2.	Although a CV outcome trial with tirzepatide has not been completed it is in progress, even though the FDA no longer mandates such trials. Analysis of previous CVOTs has demonstrated that improvements in biomarkers were very good predictors of CVOT results. This has been determined by mediation analysis of the LEADER trial (Buse et al) and utilization of a more sophisticated modern risk engine in	The revised model base case is using the HR and its uncertainty from SURPASS-4. We believe this represents the best currently available evidence, and acknowledge that uncertainty remains, especially around the addition of GIP inhibition.

	other trials. Thus, it is very likely that a CVOT will have a beneficial outcome and such a set of results can be fitted into the model for determination of cost-effectiveness.	

#	Comment	ICER Response
Patient/Patient Groups		
Partnership to Improve Patient Care (PIPC)		
1.	ICER's assessment is, once again, premature. With this report, ICER continues its concerning habit of undertaking assessments at far too early a date to have accurate inputs for its assessment. In this case, ICER has undertaken this exercise before final results on primary outcomes, such as the relative risk of major cardiovascular and renal events, are available from phase three trials. We are troubled with ICER's release of consistently more premature reports. Payers are clear that they use ICER's assessments in their decision-making processes, and PIPC has real concern that this assessment will lead to negative impacts for patients when it is based on incomplete data.	With this public comment, PIPC continues its counterintuitive arguments that our health systems, prescribers, and patients with type 2 diabetes will be able to make sound decisions around newly approved treatments without complete data and without independent treatment assessments that are discussed in the public domain. We are a broken record when it comes to recognizing that there are often data gaps for treatments that receive regulatory approval. We hypothesize that any independent patient advocacy group would only amplify such concerns around critical data gaps for newly approved treatments. Patients deserve better. The harsh reality is that patients, clinicians, and insurers are faced with difficult decisions about how best to use and pay for these new agents once approved for use. The field of health technology assessment and many professional societies view comparative clinical effectiveness research, and cost-effectiveness modeling as useful and important ways to identify the key inputs that impact the effectiveness and fair price of a newly approved therapy. Finally, we remind PIPC that we interviewed patients with T2DM prior to drafting this report. Patient experiences and the status quo motivated this draft report where we advocate for high value care for all Americans. It remains puzzling to us as to why we are not more aligned with an organization that has "improve patient care," in their name.

2.	There was no differentiation between therapies with respect to adherence rates in the model ICER chose not to investigate difference in adherence rates for different therapies. This is concerning as one of the key potential value drivers for a chronic and progressive disease, like diabetes, is the role of relative adherence to treatment. Recent studies have suggested the hospitalization and mortality rates can be twice as high in non-adherent patients than in adherent patients.	We appreciate that overall adherence to diabetes treatment remains challenging, a theme we heard from our interviews with patients, and we highlighted this in our report (see the Patient and Caregiver Perspectives section of the report). We have added language highlighting the consequences of non-adherence to medications in that section. The discontinuation rate in the model was derived from the EMPA-REG-EXTEND trial, the only trial to present discontinuation data contingent on a successful initial treatment period. We were unable to differentiate between treatments, the discontinuation rates contingent on successful treatment due to lack of mature data availability.
3.	The ICER model omits outcomes that matter to patients. ICER does not incorporate the benefits of weight loss and achievement of glycemic control, two factors patients highly value, in its report. A large majority of Type 2 Diabetes patients both globally and specifically in the United States are overweight or obese. Obesity is known to complicate their disease and worsen outcomes in those patients. A major potential benefit of tirzepatide is its impact on obesity and the complications that stem from obesity.	Weight loss and glycemic control are important to patients, and we included them in our comparative value and comparative effectiveness sections of our report. The model addresses obesity in its predictions to the extent that BMI is a predictor in several of the UKPDS event equations, including all-cause mortality. Unfortunately, we did not have strong enough data to run a subgroup analysis as a scenario.
4.	ICER's model oversimplifies the disease and fails to capture full benefit to patients. ICER's model assumes the only quality of life effects generated by a new therapy are movement between broad health states. As we have discussed in past comments, the reality for patients is the incremental improvements matter deeply, and improvements in one area can lead to other benefits, like increased productivity or reduced anxiety that make a significant positive different in patients' lives.	One of the key features of our model is that it is a patient-level microsimulation rather than a cohort model. So, contrary to the comment, all events are possible in each model cycle simultaneously. And history of events can influence future events. Additionally, we did perform societal perspective calculations and present that scenario in the supplement. There is surprisingly limited evidence on how treatments like tirzepatide may impact work productivity. We suggest this as an area for future research.

5.	The costs for cardiovascular and renal hospitalization events in the model are based on a younger population than the population of need. This is likely to underestimate true cost savings from effective treatment. Thee cost estimates used in the model specific to major cardiac events and disease sequelae are taken from a study that was limited to patients under the age of 65 The description of the ICER model very clearly states that each patient simulation is run for a lifetime, so the majority of the time for which ICER' is modeling, the patients are over the age of 65. There is considerable evidence in the literature that costs associated with hospitalization for both cardiovascular and renal events for patients over 65 years of age are significantly higher than for those patients younger than 65 years of age. Since costs are shown to increase with age, it is likely the costs ICER uses in the model are underestimates, which means any cost savings from reducing risks of events due to successful treatment will be underestimated in the model. We encourage ICER to update the inputs for the cost data, so the model more accurately captures the cost savings.	We believe that the cost inputs are reasonable, especially given that commercial payers typically reimburse in the range of 2-3x Medicare's reimbursement rates. Therefore, if anything, our analysis may give a greater cost savings to a treatment that is associated with fewer events compared to a pure Medicare perspective.
6.	ICER continues to rely on the discriminatory QALY. PIPC would like to reiterate the point it has made to ICER in past comment letters that the use of the Quality-Adjusted Life Year (QALY) is inappropriate in assessing treatments for chronic illnesses. The QALY is known to discriminate against those with disabilities and chronic illnesses, like type 2 diabetes. We encourage ICER to look to more innovative methods to assess value that do not immediately put treatments for those with disabilities and chronic illnesses at a disadvantage.	We appreciate the concerns about relying solely on QALYs. They are not used in the assessment of the comparative net health benefit: see Figure 3.1 for more details on the ICER Evidence Rating Matrix. They are also only one component of the value assessment.